KERA UPDATE

February 2000 | *About The Kentucky Education Reform Act* | #42

The Troubling Situation With The 1998 National Assessment of Educational Progress (NAEP) 4th Grade Reading Assessment

When the *1998 NAEP Reading Report Card for the Nation and The States* [1] was released on March 4th, 1999, a problem with the state-level data was immediately apparent. Some states experienced a considerable increase in the number of students with learning disabilities (SD) who were excluded from testing due to conflicts between requirements in their individual education plans (IEP) and NAEP testing guidelines. Apparently, provisions in the Individuals with Disabilities Education Act of 1997 (IDEA) made it impossible to conduct uncorrupted and meaningful tests of many SD for printed text reading ability. Because of IDEA, many more children were excluded from the NAEP in 1998. The higher exclusion rates raised questions about whether some state level NAEP scores accurately reflected real performance and could be fairly compared to other states.

Kentucky offers one of the most dramatic examples of the problem. Kentucky's six point score increase between 1994 and 1998 was one of the best improvements of any state. But, it was matched by an equally large increase in the percentage of students who were excluded from taking the test (up from 4% to 10%). All of these excluded students were SD. The 1994 to 1998 change contrasts sharply with the state's change in scores and rates of exclusion between 1992 and 1994. In addition, from 1992 to 1998 the state's total SD population skyrocketed from 7 to 13 percent of the raw sample, an 86 percent increase. That moved the state from two points below to two points above the national average which also increased in this interval.

The change in Kentucky's NAEP sample may be a bit easier to grasp using Figure 1. Notice that the NAEP 4th Grade reading assessments in both 1992 and 1994 tested a total of 96 percent of the raw Kentucky sample. Only 4 percent of the raw sample was excluded in both years. That changed very dramatically in 1998 when 10 percent of the raw sample was excluded. The increase was so large that some students who would have been classified in prior years as completely non-disabled were also excluded as SD in 1998.

Clearly, the 1998 NAEP totally ignored a significant part of the most educationally challenged in Kentucky's student population. Ominously, the group of students labeled SD was growing rapidly.

The foregoing made it seem fairly obvious that Kentucky's 1998 NAEP score might have risen simply because many more weak students were not allowed to participate. The question was, what was the real impact of the increased exclusion on the NAEP scores?

For an answer, the National Center for Educational Statistics (NCES) first turned to the experts at Educational Testing Service (ETS). ETS creates NAEP and provides technical administration and scoring services, too. A preliminary ETS report was issued in memo format by NCES on May 13, 1999 [4].

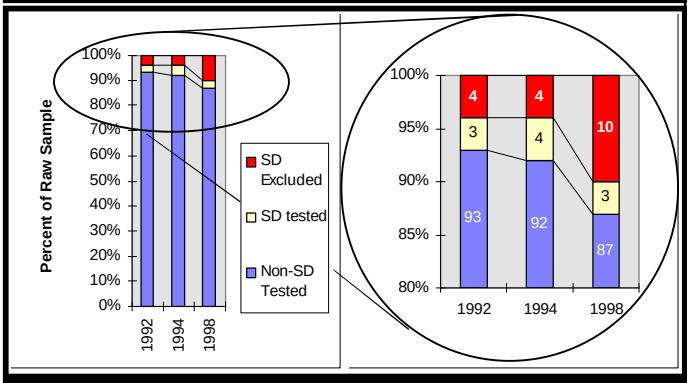
The ETS Memo has a list of questions posed by NCES along with initial answers to those questions. The most critical

Table 1
Kentucky's NAEP 4th Grade Reading Scores, Exclusion Rates for Students with Disabilities As Percent of Raw NAEP Sample, And Total Percent of Students with Disabilities As Percent of Raw NAEP Sample, By Year

	1992	1994	1998
NAEP Score	213	212	218
SD Exclusion Rate	4%	4%	10%
Total SD, Percent	7%	8%	13%

Data Sources: ([2]), ([3] Pg. 108), ([1] Pgs. 113, 163)

Figure 1
Breakdown of Kentucky's Raw NAEP 4th Grade Reading Sample



Note: The term IEP rather than SD was used to describe students with disabilities in 1992 and 1994.

question of all is number 4. In the memo, this reads:

> ***4. How would gains in State NAEP reading scores have been affected if exclusion rates had been equal across years?***

The highly significant answer from ETS:

> ***The real answer to this question can never be known.***

ETS lists a number of reasons why this is so, but the basic point is that data collected during the 1998 NAEP reading assessment is insufficient to answer this absolutely crucial question. That's because the excluded students were not tested in any way on NAEP, so there is no way to know how they would score.

Although they said that an exact answer will remain a mystery, ETS did engage in some rather controversial "what-if" analysis to try to estimate the scoring error that had been introduced by the increase in exclusions. The ETS analysis indicated possibilities ranging from no statistically significant improvement for Kentucky to a statistically valid rise in the state's score. However, ETS did not publish the possible range of scores from their analysis, and ETS did not defend the full, six point rise in the Kentucky scores between 1994 and 1998, either. As a side comment, it is easy to show that other, entirely plausible assumptions about the 1998 NAEP provide far more pessimistic results than the ETS scenarios (See Attachment 1 for examples).

The ETS report was a tremendous disappointment to the Kentucky Commissioner of Education. Eager for evidence that Kentucky's decade old reform had improved educational performance, he requested another study. Not too surprisingly, considering Kentucky's Commissioner sits on the board that governs the NAEP, he got his request.

This second study was performed by Dr. Lauress Wise of the Human Resources Research Organization (HumRRO). For several years, HumRRO has been a contractor to the Kentucky Department of Education to conduct research on Kentucky's state run assessment. Dr. Wise's report, released on September 27, 1999, asserts that Kentucky made statistically significant improvement on the NAEP and that the impact of the exclusions was almost negligible [5].

It is important to examine how Dr. Wise generates this result which differs quite sharply from the ETS findings. Wise develops 'equivalent' NAEP scores for Kentucky's excluded SD by using their test results from a state-wide assessment used only in Kentucky, the Kentucky Instructional Results Information System (KIRIS). There are a number of questions about the validity of this approach.

1. Dr. Wise's report hinges on a very critical question. **Is it appropriate and valid to consider the Kentucky 4th grade 'reading' assessment to be comparable to NAEP, especially for SD?** Wise assumes that the answer is yes. But, there was plenty of information available to Dr. Wise and NCES that this basic assumption was highly problematic. However, the Wise report offers no defense of its crucial assumption that the NAEP and KIRIS reading assessments are comparable, and Wise never even mentions this critical validity issue.

Both Wise and NCES should have known better. I cautioned the NCES Associate Commissioner for Education Assessment in June, 1999, that there were disturbing questions about the manner in which the KIRIS 4th grade reading assessment was actually administered to the SD. A discussion of these problems is found in two reports on Kentucky's SD performance on KIRIS by RAND researcher Dr. Dan Koretz. RAND data for the years 1995 and 1997 (see Table 2)

Table 2.
Percentage of Students With Disabilities Receiving KIRIS

Accommodation	1995	1997
None	19	19
Oral Presentation	**72**	**72**
Paraphrasing	49	48
Dictation	50	55
Cueing	10	10
Technological aid	3	34
Interpreter	2	1
Other	8	9

Sources: 1995 Data ([6], Pg. 13)
1997 Data, ([7], Pg. 12)

indicates that during this period about 3 out of 4 Kentucky SD consistently received an "oral presentation" on KIRIS [6], [7]. In other words, the Kentucky 'reading' assessment was probably read to most of these students. That amazing situation is actually allowed in Kentucky if SD have a reading accommodation listed in their IEP. Not only does it looks like massive numbers of Kentucky SD had their KIRIS tests read to them by proctors, but many SD also received at least one of the other KIRIS accommodations in Table 2 as well. Most or all of these other accommodations are probably incompatible with the NAEP 4th grade reading assessment rules, too.

A second piece of evidence strongly reinforces the conclusions from the RAND study. This evidence comes from the NAEP itself. Using NAEP data in Table 1, it can be calculated that 10/13, or 77 percent, of Kentucky's total SD cohort was excluded from 1998 NAEP testing. That is remarkably consistent with the percentages of the SD population that Koretz found had the reading accommodation on KIRIS in both 1995 and 1997 (72%). In addition, discussions with NAEP technical experts indicate that the presence of a reading accommodation in a student's IEP was automatic grounds for exclusion in the 1998 NAEP 4th Grade Reading Assessment.

Taken together, unless an unknown demographic shift occurred between 1997 and 1998, the data from Koretz and NAEP makes it highly likely that almost all of the Kentucky children excluded from NAEP in 1998 had the KIRIS 'reading' test read to them. If so, these kids did not take a real reading test in KIRIS. They actually took a spoken language comprehension test.

Thus, there is a very strong possibility that the 1998 KIRIS results don't tell us <u>anything</u> about whether the Kentucky SD excluded from NAEP can read printed text. If this is actually correct, it would be misrepresentation of the highest order to compare the Kentucky SD's 'spoken word' KIRIS scores to real reading results for kids who actually took the NAEP.

Sadly, both Wise and NCES are silent on the critical issue of whether NAEP's excluded students with learning

disabilities really were evaluated for printed text reading in KIRIS. In light of the evidence outlined above, that is a very unfortunate and serious omission. Absent proof that the excluded students took a real printed text assessment with KIRIS, Dr. Wise's report must be considered very incomplete and essentially of no value. The present situation also makes NCES acceptance of the Wise report as the final word on this matter highly inappropriate.

2. Aside from potentially fatal problems with its basic assumption, Dr. Wise's report demonstrates some highly questionable analysis. Some Kentuckians with mild learning disabilities could comply with NAEP guidelines and did take the 1998 NAEP on their own. Wise indicates their average score was 176.7 on NAEP's 500 point scale ([5], Table 1). But, when Wise converts KIRIS scores for the excluded kids to NAEP equivalents, the averages work out to something between 200.1 and 206.5 ([5], Table 2). This is summarized in Table 3.

Table 3.
Wise's Calculated Scores for Kentucky 4th Graders with Learning Disabilities Who Did Take the 1998 NAEP and for His Two Models for Those Who Were Excluded

Average Score for SD Who Took NAEP Unaided	176.7
Average 'NAEP Equivalent' Score for Dr. Wise's Model 1	200.1
Average 'NAEP Equivalent' Score for Dr. Wise's Model 2	206.5

Source: Table 1, Subsample 2, Stratum B, and Table 2, NAEP Equivalent, Model 1 and Model 2 [5]

The scores in Table 3 lead to a rather amazing conclusion: I**f we accept Wise's findings, we absolutely have to accept the notion that Kentucky's strongest students with disabilities, those who could read on their own and take NAEP unaided, were *significantly* outscored by other students with more severe learning problems who quite possibly might not be able to read printed text at all.** Put bluntly, this is incredibly difficult to accept. This score 'inversion' certainly adds more weight to the contention that the Wise report is fundamentally flawed.

Dr. Wise doesn't discuss his amazing inverse scoring results, raising more questions about thoroughness.

3. Unfortunately a rather disturbing conclusion does follow from the data in Table 3. NCES published scores for Kentucky SD who took NAEP 4th Grade Reading in 1992 and 1994 when more Kentucky SD took the NAEP (Attach. 2). And, as shown above, Dr. Wise calculated the NAEP score for those SD who actually took the assessment in 1998. These are all listed for comparison in Table 4.

Table 4
NAEP Scores for Kentucky Students With Disabilities, by Year

Year	Kentucky SD NAEP Score
1992	185
1994	168
1998	176.7

Sources: 1992 and 1994, (Attach 2)
1998, ([5], Table 1)

Table 4 shows that the SD tested in 1998 did not score nearly as well as the 1992 SD. But, the 1992 SD group would reasonably be expected to include many more weak students than was true in 1998. That is because all but 4 percent of the students were tested in 1992, while 10 percent were excluded in 1998 (recall the discussions about Figure 1). **So, comparing Dr. Wise's calculated 1998 SD score to data for 1992 indicates Kentucky isn't being successful with SD.** That finding is very different from the flavor of Dr. Wise's report.

By the way, inspection of Table 5 to the right indicates that **Kentucky is far from alone in its disturbing performance with students with learning disabilities** (Also see Attachment 2)**.**

Table 5 shows changes in NAEP scores for states that took the NAEP in 1992 and at least one more time thereafter. Table 5 only shows states that had a statistically valid SD sample in both years.

Virtually every state in Table 5 shows declines in scores for their learning disabled population. Only those scores shown in black background, just one state per each year grouping, indicate improvement.

Unfortunately, a number of states had so many SD excluded in 1998 that their remaining SD samples were very small.

Table 5
Summary of Changes in Scores For SD, by Year and State

	1992 to 1994		1992 to 1998	
	State	Difference	State	Difference
1	AL	-12	AL	-10
2	AR	-20	AR	-25
3	AZ	-11	AZ	-10
4	CA	-22	CO	-13
5	CO	-35	DE	-9
6	CT	-9	FL	-29
7	DE	-28	HI	-30
8	FL	-9	IA	-13
9	GA	-26	MA	-20
10	HI	-37	ME	-4
11	IA	-18	MN	-16
12	KY	-17	MO	-9
13	LA	-13	NH	-20
14	MA	-13	NM	-24
15	MD	-18	RI	-13
16	ME	-15	SC	-2
17	MN	-24	TN	-13
18	MO	-29	TX	5
19	NC	-6	UT	-21
20	NH	-13	VA	-11
21	NM	-15	WY	-13
22	NY	-11		
23	RI	-12		
24	SC	-24		
25	TN	-17		
26	TX	-3		
27	UT	-26		
28	VA	1		
29	WI	-14		
30	WV	-23		
31	WY	-14		

(Source: [8], Pgs. 63-71)

Thus, NCES didn't report 1998 SD scores for these states. Aside from Kentucky, states with missing 1998 SD scores include several with highly aggressive education reforms such as Maryland and North Carolina. In addition, the state with the overall best NAEP improvement from 1994 to 1998, Connecticut, also had an insufficient SD sample in 1998 and did not receive SD scores. Connecticut did have a 9 point decline in SD 4th grade reading scores between 1992 and 1994.

The point here is that while NAEP may not provide accurate information about whether these states are making progress, it does offer disturbing clues that increasing

educational failure with SD could be hiding behind rapidly increasing exclusion of the learning disabled from meaningful assessment participation.

This situation makes it very inappropriate to gloss over what is happening with the growing number of students with learning disabilities in Kentucky and elsewhere. Unfortunately, the Wise report and its uncritical acceptance by NCES act to provide exactly such a glossing over of a very serious situation.

4. One other technical point: The validity of KIRIS became so suspect in Kentucky that this assessment was totally abandoned after 1998. Beginning in 1999, a new and quite different Kentucky assessment was launched which totally ignores the old KIRIS scores. **Thus, Dr. Wise's study is totally predicated on results from an obsolete assessment that was abandoned for cause.** Dr. Wise never mentions this important point in his study, although, as a contractor to the Kentucky Department of Education doing validity research on the Kentucky assessment, Wise had to be fully knowledgeable of this situation.

Other Notes

NAEP elementary school reading tests deal with children who are more than half way through their primary school years. **The idea that growing numbers of 4th grade children in any state have not been taught to read well enough to cope with a reading assessment is very disturbing, especially when the NAEP data shows we accomplished this task better with a much larger percentage of our learning challenged children in the early years of the decade. It is difficult not to believe that the rapid increase in the number of students in some states who are being labeled as learning disabled and saddled with test-corrupting accommodations is more an indication of school failure rather than a result of a real shift in the demographic makeup of school populations.** To be blunt, if the 86 percent rise in students with learning problems in Kentucky between 1992 and 1998 is real, then the Centers for Disease Control and many other agencies should descend upon the Bluegrass State in droves to find the cause of this epidemic of mental deterioration.

Closing Comments

At best, this entire matter is rather extraordinary. Issues surrounding the 1998 NAEP have already been the subject of one quite heated Congressional inquiry. While that inquiry focused on political issues, which are most certainly troubling, the Congress never discussed any of the equally disturbing NAEP SD sampling problems outlined above.

It must also be noted that the former federal Commissioner of Education Statistics resigned recently under a cloud of political questions. Dr. Wise was nominated as his replacement in February, 2000.

Hopefully, this report will stir the Congress to reopen this matter. We need to find out if NAEP is being corrupted to an unacceptable level by exclusion of students with learning disabilities. We need to determine whether Dr. Wise's study can withstand close scrutiny by the technical education community and meets standards of thoroughness and accuracy that we would want in a new Commissioner of Education Statistics. And, we need to know if the increasing trend in students tagged as learning disabled is going to continue. With far too many critical decisions concerning radical education reform hanging in the balance, answers simply must be found.

Most certainly, the controversial impacts of the Individuals with Disabilities Education Act of 1997 need revisiting. As things stand, IDEA forms a barrier both at the federal and state level that separates parents, the general public, and state and federal leaders from the truth about school performance. Without modification, IDEA apparently creates a climate that is absolutely hostile for accountability assessment programs. And, coupled with current accountability trends, IDEA seems to create powerful pressure to label children as learning disabled when that may not really be the case.

Bibliography

[1] Patricia Donahue Et. Al., *NAEP 1998 Reading Report Card for the Nation and the States*, National Center for Educational Statistics Report NCES 1999-500, Washington, DC, March 1999. Downloadable from NCES web site at http://nces.ed.gov

[2] *Table B.7, NAEP 1992 Reading Report Card for the Nation and the States*, Part of Faxed letter from Lawrence Feinberg, National Assessment Governing Board, Washington, DC, 17 May 1999.

[3] Jay R. Campbell Et. Al., *NAEP 1994 Reading Report Card for the Nation and the States*, National Center for Educational Statistics, Washington, DC, January 1996.

[4] John Masseo, John Donoghue and Catherine Hombo, *A Summary of Initial Analysis of 1998 NAEP Exclusion Rates*, (A published memo), Educational Testing Service, Princeton, NJ, 13 May 1999. Online at: http://nces.ed.gov/pressrelease/naep599ets.html

[5] Dr. Lauress Wise, *Impact of Exclusion Rates on NAEP 1994 to 1998 Grade 4 Reading Gains in Kentucky*, National Center for Education Statistics, Washington, DC, 27 September 1999. URL: http://nces.ed.gov/commissioner/ remarks99/9_27_99.asp#approach

[6] Daniel Koretz, *The Assessment of Students with Disabilities in Kentucky, CSE Technical Report 431*, National Center for Research on Evaluation, Standards, and Student Testing, UCLA, Los Angeles, CA, July 1997.

[7] Daniel Koretz and Laura Hamilton, *Assessing Students with Disabilities in Kentucky: The Effects of Accommodations, Format, and Subject, CSE Technical Report 498*, National Center for Research on Evaluation, Standards, and Student Testing, UCLA, Los Angeles, CA, January 1999.

[8] *NAEP 1998, 1994 and 1992 National and State Reading Summary Data Tables for Grade 4 Student Data, Weighted Percentages and Average Composite Scale Scores*, National Center for Educational Statistics, Washington, DC, 19 February 1999. Downloadable from NCES web site at http://nces.ed.gov.

ATTACHMENT ONE

Alternate Analysis of Kentucky's NAEP Exclusion Impacts

1. One Common-Sense Way to Analyze the NAEP Exclusions

As discussed in the main paper, ten percent of NAEP's initially selected student pool for Kentucky was not tested in 1998. It is very likely that many, perhaps virtually all, of these students were excluded because their Individual Education Plans (IEP) required all questions on their tests to be read by proctors.

Clearly, a requirement to read all questions to students is totally antagonistic to the goal of determining if these students can read. Also, when teachers decide that 4th grade students require reading accommodations on all tests, these teachers have essentially declared that those children cannot read. Thus, a not unreasonable way to analyze real reading performance in Kentucky might be to score all such children with a zero and average those scores with the average score for the rest of the students (218 on the 1998 test). If that is done, the following results:

$$\text{Corrected Reading Score} = \frac{(218 \times 90\%) + (0 \times 10\%)}{100\%} = 196$$

That, of course is a tremendous drop from 1992 and 1994.

Some might argue that because Kentucky excluded 4 percent of students in the earlier assessments, that this should be properly considered. A way to do that is to weight the average of only the additional 6 percent that got excluded in 1998. Doing this gives

$$\text{Score, 6\% Exclusion Increase Corrected} = \frac{(218) \times 90\% + (0 \times 6\%)}{96\%} = 204$$

This is still a major, statistically significant, drop from the 1994 score of 212.

If Kentucky students who were eliminated from 1998 NAEP testing are scored with a 0 for reading (which certainly would be warranted assuming they had to have all questions on tests read to them) and if those scores were averaged in with the average score for the students who actually tested, the state's 1998 NAEP score would be at least 8 points lower than in 1994.

2. Linear Regression Analysis of the changes in NAEP 4th Grade Reading Scores from 1994 to 1998 Versus the Changes In Percent of Students Excluded Due to Individual Education Plan Conflicts.

Table 1, Attachment 1, below shows the change in exclusion rates and the change in NAEP 4th grade reading scores for those states that participated in both 1994 and 1998.

Table 1, Attachment 1

State	Change in Students Excluded as a Percent Of Total NAEP Sample	Change in NAEP Reading Score	Averages For Different Exclusion Rates
LA	7	7	**7**
KY	6	6	**6**
NC	5	3	**5**
SC	5	7	
CT	4	10	**5**
IA	4	0	
AL	3	3	
NM	3	1	**2.3**
WV	3	3	
MA	2	2	
MD	2	5	
MO	2	-1	**1.8**
RI	2	-2	
VA	2	5	
DE	1	6	
GA	1	3	
MT	1	4	**4.4**
TX	1	5	
WA	1	4	
AZ	0	1	
HI	0	-1	
NY	0	4	**0.4**
WI	0	0	
WY	0	-2	
CO	-1	9	
MN	-1	4	**5.3**
NH	-1	3	
AR	-2	0	
CA	-2	5	
FL	-2	2	
ME	-2	-3	**0.43**
MS	-2	2	
TN	-2	-1	
UT	-2	-2	

Sources: NAEP 1998 Reading Report Card for the Nation and the States, Pages 113 and 163
NAEP 1994 Reading Report Card for the Nation and the States,

Notes for Table 1, Attachment One:
Correlation of Change in Exclusion to Increase in scores = 0.41307

A scatter plot of the data with a linear regression line appears in Figure 1, Attachment 1.

A couple of observations are possible.

First, the slope of the regression line, 0.54, implies that for each one percent increase in SD exclusions, there was approximately a half a point increase in score. For Kentucky's 6 point increase in exclusion, that would mean 6 times 0.54 or an error of about 3.2 points due to the effect of exclusion.

The Y intercept of the regression line (2.11) implies the overall average improvement for all states on NAEP 4th grade reading between 1994 and 1998 was closer to 2 points rather than the 3 points actually posted. That isn't a terribly strong improvement on a 500 point scale test, especially since most of it can be explained by statistical sampling error alone. It is also a 1 point drop from 1992. This raises questions about the possible corruption of scores for other states besides Kentucky.

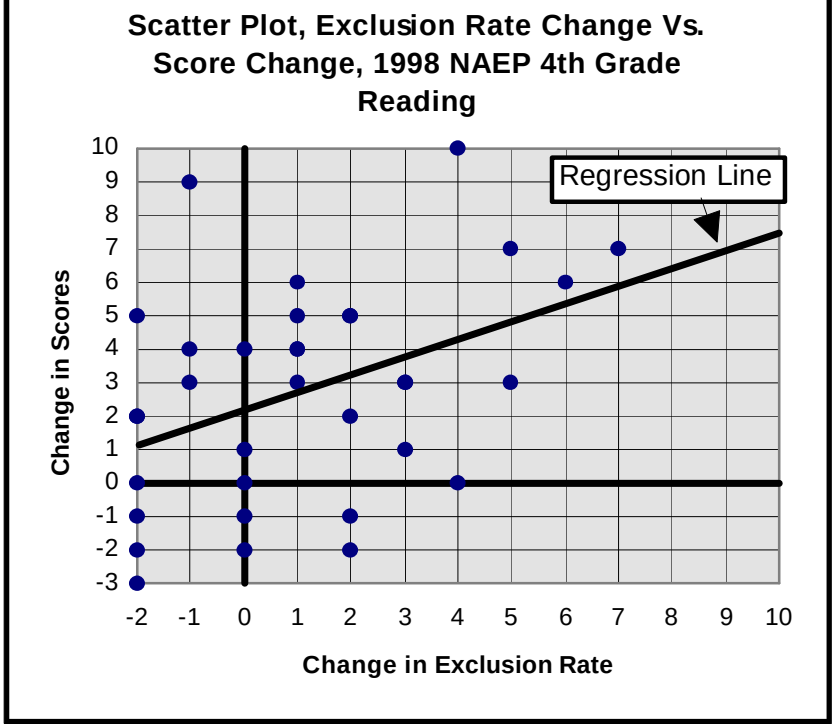
One problem with linear regression is that it assumes a straight-line relationship exists between the data for all points in the database. Consider the data on the right side of Table 1 in this attachment. Here you will see score averages **(In Bold Typeface)** for states arranged according to their changes in the rate of IEP exclusions from 1994 to 1998. Notice that as the change in NAEP exclusion rises above 2%, the score average also rises, and by increasing amounts. This implies the linear model isn't accurate across the entire spectrum of changed exclusions.

Potential non-linearity was explored by doing a piecewise regression analysis of only those states that had an exclusion change of plus 2 percent or more from 1994 to 1998. The slope of this piecewise line is 1.1, with a Y intercept of - 0.43. That implies virtually all of the score increase for states at the top of the listing in Table 1 represents no real improvement between 1994 to 1998. Real performance for these states may have even declined a bit. As a result, it seems fair to say that:

Regression analysis of published NAEP 4th grade reading score changes and exclusion rate changes indicates no less than 3.2 points of Kentucky's 6 point score increase might be solely caused by increased exclusion of students with disabilities. And, piecewise linear regression analysis implies a still higher inflation occurred due to the high rate of exclusion in Kentucky. Either of these situations would mean Kentucky's score change was not statistically significant.

<u>It must be noted that this *is not* hypothetical "what-if" modeling. This regression analysis is based on hard data including actual NAEP scores and actual exclusion trends.</u>

Figure 1, Attachment 1



Slope of Regression Line = 0.54

Regression Line Intercept = 2.11

ATTACHMENT TWO

Extract from NAEP 1998, 1994 and 1992 National and State Reading Summary Data Tables for Grade 4 Student Data

The table to the right summarizes the average scores for students with disabilities who participated in NAEP 4th grade reading in the 1992, 1994 and/or 1998 administrations.

***** Cells filled with asterisks identify states that had an insufficient sample of SD to permit publishing scores. **Notice the *big* increase in the number of such states in 1998.**

The trends on Table 1, Attachment 2, show general decline in SD performance on NAEP is endemic across the country.

Because the national percentage of students classified as SD has risen, this table offers disturbing evidence that <u>1998 NAEP sampling of the IEP cohort was insufficient in many areas of the United States.</u>

Question: Does the NAEP create pressure for schools to categorize students as learning disabled in ways that will prevent these students from being tested and lowering overall scores?

Clearly, if schools can prevent SD students from being counted in the NAEP totals, it will boost the scores. But, society will not be given an accurate picture of how schools are really performing, and the excluded students will be getting schooling of totally unknown quality.

Table 1, Attachment 2
NAEP 4th Grade Reading Scores for Students with Learning Disabilities

	1992		1994		1998	
	State	SD Score	State	SD Score	State	SD Score
1	AL	185	AL	173	AL	175
2	AR	177	AR	157	AR	152
3	AZ	183	AZ	172	AZ	173
4	CA	175	CA	153	CA	*****
5	CO	192	CO	157	CO	179
6	CT	204	CT	195	CT	*****
7	DE	179	DE	151	DE	170
8	FL	192	FL	183	FL	163
9	GA	196	GA	170	GA	*****
10	HI	178	HI	141	HI	148
11	IA	193	IA	175	IA	180
12	KY	185	KY	168	KS	186
13	LA	191	LA	178	KY	*****
14	MA	215	MA	202	LA	*****
15	MD	192	MD	174	MA	195
16	ME	204	ME	189	MD	*****
17	MI	*****	MN	175	ME	200
18	MN	199	MO	171	MI	*****
19	MO	200	MS	164	MN	183
20	MS	*****	MT	176	MO	191
21	NC	182	NC	176	MS	*****
22	NH	200	NH	187	MT	*****
23	NM	189	NM	174	NC	*****
24	NY	188	NY	177	NH	180
25	OK	197	RI	186	NM	165
26	RI	198	SC	167	NV	*****
27	SC	191	TN	169	NY	*****
28	TN	186	TX	187	OK	*****
29	TX	190	UT	161	OR	171
30	UT	187	VA	200	RI	185
31	VA	199	WA	164	SC	189
32	WI	198	WI	184	TN	173
33	WV	196	WV	173	TX	195
34	WY	194	WY	180	UT	166
35					VA	188
36					WA	166
37					WI	*****
38					WV	*****
39					WY	181

Source: [8], Pgs. 63-71.
Only participating states in each administration of NAEP are shown.

Short Comments on NAEP

The National Assessment of Educational Progress (NAEP) is a federal program administered by the National Center for Education Statistics (NCES) and governed by the National Assessment Governing Board (NAGB). NCES is a part of the federal Department of Education, and the US Secretary of Education appoints all of the members of the NAGB, so this program is solidly under control of the Secretary.

There are three different assessment programs in NAEP:

1. Main National Assessment
2. Long-Term Trend National Assessment
3. State Assessment

The first two have been in use since the 1969-70 school year. The first state NAEP was given for 8th grade math in 1990.

The Main National Assessment provides US-wide average scores. It does not include any state level data. The Main Assessment changes over time so that the most recent tests can only be compared to the last two or three previous tests.

The Long-Term Trend National Assessment maintains the same testing conditions and types of questions over time. There are no state results in this program, either. Long-Term NAEP provides more accurate long-term information, but not as many subjects are covered with this program.

The state assessments are the newest NAEP. They are possibly the most important today. That's because there are many radical education reforms under way in many states, and people would like to use the NAEP to chose the best among them.

NAEP has assessed many subjects over the years including Math, Science, Reading, and Writing, for example. But, state level NAEP results are quite limited. As of the 1998 assessments, only 8th and 4th grade math and 4th grade reading had been given more than one time, so this is the only state level trend data available from NAEP.

NAEP scoring uses either scale scores, usually on a 0 to 500 scale, or achievement level scores which include just four grades of 'Below Basic,' 'Basic,' Proficient,' and 'Advanced.' Scale scores are highly massaged from the raw scores so that it is unlikely any state will ever score below 150 or above 350 on any NAEP assessment using the 500 point scale. The Achievement level results are much more problematic, and even testing experts sharply disagree on their accuracy and worth.

Complexity with NAEP scoring has led to past problems. NAEP 4th grade reading scores had to be corrected after the 1994 testing, and early state level math scores required corrections when the 1996 math scores were released.

National NAEP assessments do not report results lower than regions of the country. No state level results are available in the Main or Long-Term Trend assessments. State NAEP do not report results below state level, at least according to ***The NAEP Guide, 1999 Edition.***

NAEP uses a lot of open response (written answer) questions and also asks a lot of demographic questions on each administration. In fact, the majority of questions each child answers are demographic, not academic. Both types of questions have raised criticisms concerning both student and parent privacy. For example, questions on reading may solicit students to use "personal background" in their answers. And, demographic questions gather information about parents' education levels, for example.

The Wise report, discussed in the main part of this ***KERA Update,*** demonstrates how demographic questions create privacy issues even if the students' names are removed from NAEP datafiles. Dr. Wise achieved an "exact" match for 86 percent of Kentucky's raw student sample and their NAEP files using demographic data alone. In other words, he exactly matched 2,358 out of 2,741 Kentucky students without having to use their names. This means that anyone who gets access to the restricted NAEP datafiles and has access to student demographic files can do a similar matchup. Of course, NAEP managers are emphatic about the level of protection afforded these restricted access files. But, considering that computer hackers have recently penetrated all sorts of sensitive computer files, the fact that NAEP data can be manipulated using demographics alone poses a non-trivial threat to the privacy of students and their parents.

For a readable reference on the NAEP, check ***The NAEP Guide, 1999 Edition***, Report 2000-456, National Center for Education Statistics, Washington, DC. On-line at www.ed.gov.

Acronyms in this Report

ETS	Educational Testing Service
HumRRO	Human Resources Research Organization (Dr. Wise's organization. Contracts to Kentucky Department of Education)
IDEA	Individuals with Disabilities Education Act (Renewed in 1997)
IEP	Individual Education Plan
KIRIS	Kentucky Instructional Results Information System (The assessment program used in Kentucky between 1991-92 and 1997-98)
NAEP	National Assessment of Educational Progress
NCES	National Center for Education Statistics
RAND	A Washington-based "think tank" doing education research
SD	Students with Learning Disabilities (Called "IEP" in early NAEP Report Cards)