Is There Really Quality in *Quality Counts 2001* Ratings?

There's a tremendous amount of unscientific information concerning what really works in public education. Though the rhetoric is extensive, thoroughly convincing proof that education programs actually work is often hard to find. A few weeks ago, the *Education Week* newspaper provided a new case in point.

EdWeek recently released its annual report on education, *Quality Counts 2001. QC 2001* ranks states on a variety of indicators. Among the most important ratings: those given to each state's standards and accountability program. Unfortunately, it is extremely easy to challenge these *QC 2001* ratings. Indeed, high ratings in *QC 2001* might actually correspond to the some of the worst education performance.

Consider the data in Table 1. This table shows 1998 National Assessment of Educational Progress (NAEP) 4th Grade Reading scores for four top and four bottom rated states in *QC 2001* (a fifth bottom rated state didn't take NAEP and isn't included). Table 1 also includes the percentage of each state's raw sample that was identified as learning disabled and excluded from taking the NAEP.

Clearly, states that got top ratings in *QC 2001* were generally outscored on NAEP by states *EdWeek* rated at the bottom. But, there is more to the picture. It's obvious NAEP scores will rise if more weak students are barred from participating. The only question concerns the amount of inflation, and even experts cannot agree on that. My personal estimate, based on regression analysis of the 1998 NAEP, is that scores are inflated by somewhere between 0.5 to a full 1.0 points for every 1 percent increase in exclusion. Thus, because states on the left of Table 1 generally have higher exclusion rates than the states on the right, scores on the left of Table 1 are probably inflated more than scores on the right. If so, that would make the rating errors in *QC 2001* even more serious.

In addition, highly rated states in *QC 2001* generally identified quite a bit more of their children as too learning disabled to take the NAEP. That, by itself, is a highly contradictory situation. Why are there noticeably more learning problems in those states *QC 2001* rated highly?

NAEP isn't the only test at odds with the ratings in *QC 2001*. Recent performance on college entrance testing for those states in Table 1 that had generally equivalent participation rates on the ACT also supports a conclusion that high ratings for standards and accountability in *QC 2001* correspond to weaker education performance. See Table 2 for this data.

Actually, problematic ratings in *Quality Counts* aren't terribly surprising. The report is funded by the Pew Charitable Trusts. Pew spent millions of dollars in states like Kentucky pushing some rather controversial ideas about accountability and how it should serve as the central element of reform. So, when *QC 2001* claims its ratings are based on the "best thinking" about what works in education, one would certainly expect this "thinking" to be influenced by the financial sponsor's vision of what is "best." But, the real value of Pew's ideas still remains unclear.

In fact, if ACT and NAEP have any value, it becomes questionable whether the "best thinking" from Pew and *EdWeek* leads to good education for students. And, if thorough analysis of real student performance is crucial, we might do well to look beyond *QC 2001*.

Table 1. Top and Bottom Rated States for Education Standards and Accountability in *Quality Counts 2001,* with Their NAEP Scores and Their NAEP Exclusion Rates for Students with Learning Disabilities, 1998 NAEP 4th Grade Reading

	QC Top States				QC Bottom States			
	MD	NY	KY	NM	IA	MT	MN	RI
QC Grade	A	A	A-	A-	F	F	F	F
1998 NAEP Reading	215	216	218	206	223	226	222	218
NAEP Exclusion	9%	5%	10%	9%	8%	4%	3%	6%

Table 2. ACT Composite Scores and Graduates' Participation Rates, States from Table 1 with Generally Equivalent ACT Participation

	QC Top States		QC Bottom States		
	KY	NM	IA	MT	MN
QC Grade	A-	A-	F	F	F
2000 ACT SCORE	20.1	20.1	22.0	21.8	22.0
% TESTED	71	66	69	58	66