Practical Assessment, Research & Evaluation
A peer-reviewed electronic journal.

Volume 10 Number 15, September 2005 ISSN 1531-7714

A Review of Greene (2002) *High School Graduation Rates in the United States*

Richard P. Phelps
Nonpartisan Education Review

The "Greene Method" of calculating school "graduation rates" and the Manhattan Institute (MI) criticisms of official graduation and completion statistics are outlined and scrutinized. The methodology fails to recognize the complexity of the issue and appears to ignore the considerable efforts that have been undertaken by education statisticians to remediate the problems inherent to these types of data. The Greene method for calculating completion ratios is simulated and found to have little to no reliability. It is recommended that anyone intent on reporting valid and reliable education indicators avoid use of the Greene Method.

The public policy think tank, the Manhattan Institute (MI) has published *High School Graduation Rates in the United States* (Greene, 2002a) proposing a new method for calculating high school completion ratios. The method has been adopted to some prominent customers, including the Gates Foundation (e.g., Greene & Forster, 2003) and *Education Week* (2002). The author of the method has been introduced by the media to the nation as the country's leading expert on graduation rates (see, for example, CBS News, *CNN Presents*, & National Public Radio).

Most of the thousands in the United States who have worked with enrollment and graduation data know how problematic they are, and that they are widely misinterpreted. No matter what one thinks of the MI's new method, one must concede that the think tank's mass marketing has brought more attention to an underappreciated issue. Is that a good thing? This article attempts to answer that question.

BACKGROUND

Education statisticians–so under appreciated

Like any occupational group, education statisticians are often misunderstood. To liberally paraphrase Tolstoy, however, each occupational group is misunderstood in its own way. Being a statistician is somewhat like being a Secret Service Agent, or a building security guard. So long as you do your job, many folks consider you the moral equivalent of wallpaper. When something really bad happens, however, there you are at the scene of the crime, so it's likely to look like your fault and, suddenly, you're a celebrity...celebrity screw-up, that is.

One common misunderstanding about statisticians is that we create the statistics (rather than just collect and analyze them) and are, thus, completely responsible for them. They are "our" statistics. So, if there are any problems with them, we must have done a bad job. Or, even if we are not responsible for creating the statistics, we are responsible for publishing them, and so should be held accountable if we publish any that are misleading.

Education statistics–so misunderstood

A related misunderstanding stems from the conviction of some that statistics should not be published unless and until they are completely valid and reliable, as if such a status were common to the world of social measurement and easy to attain. It may be easy to be consistent with a measurement when one is in control of all its aspects, as one might be in walking around a wood shop with a single ruler measuring the lengths of newly cut planks. But, even alone in the wood shop measurement variation is likely to occur, depending on the angle of view, the position of the light, the steadiness of one's hands, one's fatigue or lack thereof, and so on. Collecting social statistics, and collecting them from a multitude of different individuals taking the measurements independent of one another, inevitably causes much more variation. No measurement is perfect, neither is any measure.

Another misunderstanding relates to the previous one and arises when new statistical series are initiated. No one can perfectly predict how things will turn out when an organization begins a new statistical collection. Particularly when data emanate from separate, independent sources, all new to a particular collection and its standards, there are bound to be kinks in the process and in the resulting numbers. It is not uncommon in the first several years of a statistical series to see previously published numbers revised as those kinks get worked out. There is no alternative to a tolerance for imperfection in statistical data, unless one wants to consider no data collection at all. Statistical series can get better over time, and approach perfection perhaps, but starting out perfect is probably not possible so long as humans are involved.

Some departments at the Organisation for Economic Co-operation and Development (OECD), a collector of social data on a massive scale, make it a policy to list member countries in its statistical tables even when they have turned in no data (see Organisation for Economic Co-Operation and Development, any year, 1999). Why? It is hoped that the sight of their name next to the blank entries, and alongside the names of other countries that did turn in their data on time, might shame them into being more diligent with their data submission, and, as it were, hold a place for them as if to say the statistical series will continue and their data are still expected. Over time, the series do get better and more complete.

Data collection agencies, such as the OECD or the U.S. National Center for Education Statistics (NCES) do accept data that are inaccurate. Usually, they do not do not do so knowingly, however, checking submissions as best they can and requesting corrected figures when they can determine that the originals are not accurate. But, they cannot always determine that submitted figures are not accurate, particularly when the magnitudes of those figures seem "reasonable" (e.g., similar to last year's, similar to those from similar jurisdictions).

Yet another misunderstanding arises from the naive belief of some that statistics are collectively exhaustive. That is, that one can tell a complete statistical story about education because all the numbers are available somewhere. In fact, some aspects of education are well covered by statistical collections, and others are not.

Ultimately, most statistical agencies have no power to coerce jurisdictions to fix bad numbers. In some cases, a statistical agency's role is simply to collect and report what is given them, with no expectation that they will even check the numbers for "reasonableness." Years ago, for example, I wanted to use a certain statistical series published by an international agency but, upon careful examination, decided that the numbers did not seem to be calculated in a consistent manner across

countries and also thought, given the nature of the particular measure, that it would be extremely difficult to maintain consistency across countries anyway. So, I telephoned the statistician in charge of the particular series and she recommended, bluntly, that I neither trust nor use the numbers. They had an obligation to collect the numbers in that particular series, she said, but no one in her organization trusted that they were accurate.

Having myself worked in the research division of a state education agency for several years, I have felt particularly amused by the righteous recriminations vented over dropout rates of late. Texas found itself to be especially excoriated for underestimating its dropout rate, as if it were a unique location for that crime. In part, some education researchers just wanted to pick on Texas (for obvious reasons in the year 2000) but virtually all states could be found guilty of the same crime.

I find the righteousness over the issue amusing for two reasons. One, anyone really familiar with education statistics knows dropout statistics are a mess, in part because jurisdictions have every incentive to under-report dropouts and over-report attendance. In most jurisdictions, school district revenues come from the state based directly, and often entirely, on the number of students the district claims to be teaching.

Two, it requires only a little common sense and some time pondering the issue--about how one might count a dropout--to see that it has to be a pretty fuzzy statistic in a society as open as ours. Potential dropouts typically do not show up at the school office and fill in an official "drop out" form; they just quit coming. Most dropouts themselves probably cannot pinpoint the moment in time when they became dropouts. Many get bored and quit going to school, thinking they will return after a break. Some take time off to help around the house, or with the household income, with the sincere intention of returning later. Some do return; some do not. Some return, only to leave again later. Given this kind of social dynamic, why would one expect district-reported dropout numbers to be statistically pure, even if the districts wanted to report them purely?

Finally, it is relevant to ask if we really desire to uniformly criticize school districts regarding their dropouts and non-graduates? This is not just a question of statistical methodology; it's a policy question. What if some of those students profoundly desire to be out of school and, when they are in school, are highly disruptive, ruining not only their own education, but everyone else's as well? Should we coerce those schools to entice those students to stay? What signal does that send to the students who behave themselves and genuinely wish to learn? Yes, of course, some schools are at fault, in whole or in part, for their dropouts. But, some are not. In the end, an individual student drops out of school because that individual student wants to.

Any state education agency statistician who pays attention knows that enrollment and attendance rates, and the resulting state education aid allotments, are often skewed in favor of poor districts. That is because poor districts, for a variety of reasons, tend to have more dropouts and transfers out, and more transient students in general. They are, thus, more likely to be reimbursed for students who might have enrolled in early September, but who are no longer there by March. I have heard more than a few education statisticians express some satisfaction that, despite their frustrations with the unreliability of district-reported dropout, enrollment, and attendance numbers, at least they knew that this particular bias in state aid allocations, caused by the "temporal decay" of the relevant statistics' reliability, most likely skewed funds toward the districts that needed the money the most.

HIGH SCHOOL GRADUATION RATES IN THE UNITED STATES (MANHATTAN INSTITUTE)

The Manhattan Institute report *High School Graduation Rates in the United States* (Greene, 2002) has received much attention and produced considerable effect. Some jurisdictions found to be lagging by the Manhattan Institute report have been roundly criticized by the press and some of them have attempted to defend themselves. To innocent bystanders, however, their defenses probably sound,

well... defensive. The Buckeye Institute applied their method to an analysis of Ohio school districts (Greene & Hall, 2002c). The Bill and Melinda Gates Foundation contracted with the Manhattan Institute (MI) to work its magic in Washington State (Greene, 2002b). The Black Alliance for Educational Options sponsored the study reviewed here (Greene, 2002a). MI's graduation rates have even now been adopted by *Education Week's Quality Counts*. It is not easy for an ordinary state- or district-level bureaucrat to compete against celebrity research with ready access to a nationwide audience.

The "Greene Method"

The Manhattan Institute report describes a method for calculating completion ratios, called the "Greene Method" (the report author's name is Jay P. Greene). Strictly speaking, the author does not calculate graduation "rates," because he employs aggregate figures from different populations in his numerator and denominator (see Appendix A, "Rates versus Ratios"). But, perhaps that is a fussy statistician nit pik. Few should quibble with his terminology if he has, in fact, happened upon a method for calculating a superior, more accurate, and more useful completion ratio.

His algorithm is this: the reported number of graduates in 12th grade *divided by* 8th-grade enrollment four years earlier in the same district or state. How does he account for student migration? He adjusts 8th-grade enrollment thus: "Actual 8th grade enrollment + (actual 8th grade enrollment x percentage change in total or ethnic sub-group enrollment in the jurisdiction" in those four years).

As a kind of stand-alone statistic for use in each jurisdiction by each jurisdiction, this might not be so bad. Unfortunately, Greene advocates using his measure to compare jurisdictions, assuring us that it is consistently well-behaved for such purposes. *["The graduation rates provided here provide simple, straightforward, and accurate information about schools nationally." (p.9)]* Further, he suggests that his statistic is appropriate to use in judging each jurisdiction's relative performance. *["The rates at which students graduate high school provide us with information about the effectiveness of those schools." (p.9)]*

Greene describes his algorithm as "remarkably simple" and, on that point, he's correct. Which begs the question: if it is so simple to calculate what Greene calls "quite accurate" graduation rates, why haven't the stalwart folks at the National Center for Education Statistics done it? Are NCES statisticians really so negligent, or so dense?

Greene makes it clear that he regards the most commonly used graduation and dropout numbers to be bogus (for good reason with respect to the latter) and their use, he asserts, is an appalling scandal (he may be right there, too). Then he calculates his "straightforward and accurate" figures based entirely on enrollment numbers that, naturally, must be as accurate and reliable as an atomic clock.

Or, are they? Enrollment numbers are not as reliable as Greene seems to think, but they are still probably fairly comparable across jurisdictions (see Appendix B, "A Brief Primer on Student Headcounts").

Of all the aforementioned statistics–for enrollment, attendance, dropouts, and graduates-graduates might be the most trustworthy. In most jurisdictions, there is no direct financial incentive to either under or over report the number of graduates. Graduates must be issued diplomas, which is a deliberate act unlikely to be carried out for a fictitious being. Many would argue that there exist strong incentives to graduate students who do not deserve to graduate, so as to avoid visits from angry parents (or students), lawsuits, and the like, but that is a completely different issue. If a student is reported to be a graduate, there is little reason to suspect he is not. Again, this is not to say that all graduations are well deserved, just that they are probably fairly accurate counts.

Why is Greene, so critical of previously reported dropout and graduation rates, so blindly trusting of the enrollment numbers? I can't answer that.

But, the "Greene Method" is absolutely reliant on their assumed unblemished veracity. To justify his work, Greene heaps doubt after doubt upon graduation and dropout rates. But, for his

recommended fix to be valid and reliable, enrollment rates must be as pure as the graduation and dropout rates are tainted, despite the fact that all three measures emanate from the same sources and are collected in the same manner.

Unfortunately, while enrollment is probably *measured* in a roughly consistent manner across jurisdictions—making it a fairly reliable indicator—it is not really an appropriate indicator to use for Greene's purpose. Given how Greene is using it, enrollment is not a *valid* indicator, as shall become apparent in the discussion below.

Attendance, by contrast, might be a more valid indicator to use, in theory, but it is wildly unreliable for comparisons across states. Its definitions, reporting requirements, reporting dates, effort, and degree of transparency vary far too much.

Time and again throughout the Manhattan Institute report, the author criticizes the past behavior of education statisticians, or the product of their efforts, even while he effortlessly brushes off the fatal flaws in his own method as minor or unimportant. For example:

["I chose to use 8th grade enrollments because some students drop out of school before 9th grade. In addition, 9th grade is a common grade in which students repeat the grade, which can artificially inflate 9th grade enrollments and understate the true graduation rate."]

There are at least two things wrong with the intention expressed above. First, some students drop out of school before 8th grade, too, and some others must repeat it, more than 10 percent of grade-level students in some states.

Second, unlike 9th grade which, in some jurisdictions, is part of high school, 8th grade is part of high school almost nowhere. So, there's even more moving around–that which occurs during the not quite uniform transfer of middle- and junior high-school students to high schools–in between the point in time of Greene's denominator and that of his numerator. Greene decides that the movement cannot be significant because he does not find in the aggregate, in the school districts where he looks, much change in the size of school populations between these two levels of education, nor even between private and public school enrollments between the two levels. In the aggregate, all students seem to remain neatly on their tracks, steadily progressing up the grade levels in their proper jurisdictions.

Of course, in and out student migration could each occur in massive proportion but, so long as the two are of approximately equal magnitude, one would not be able to detect their perturbation in the aggregate. Even so, Greene's ratio might be fine for some purposes. But, remember, he claims that his measure is consistent across jurisdictions, and suggests we use it to judge their performance. There may exist a school district where the students in 8th grade are the same and the only students who could be there in 12th grade, say in the most remote corner of Alaska. Far more common are school districts where not even a majority of the 8th graders remain four years later.

Greene adjusts for the perturbation using changes in *total* enrollment in each jurisdiction over the 4-year period which, of course, includes changes in enrollments in grades K through seven. He mentions the possibility (in a footnote) that demographic trends unique to the earlier grades could skew his numbers for the later grades, but then quickly dismisses the idea, as his method of adjustment "is the most parsimonious assumption for an adjustment and it is still likely to be reasonably accurate." But, how could he possibly know it is "reasonably accurate?" What can he benchmark against in order to check?

He goes on to argue that because of the demographic trend problem, "the graduation rate will be slightly underestimated." [In reality, it could be over- or underestimated, it all depends on the trend, and it will not necessarily be only "slight."] "However, the total student population change can also be influenced by a high rate of dropouts that could cause the graduation rate to be overestimated. In sum, there is little reason to expect systematic bias from this adjustment and it is likely that any errors are small."

Again, how could he know? …and, how convenient. In and out migration magically cancel

each other out and, even if they don't, any demographic trend that might bias his method of adjusting for that problem is neatly canceled out by respondent bias in dropout statistic reporting. Furthermore, it probably goes without saying, all this balancing out is surely perfectly consistent across jurisdictions.....

Is NCES negligent or misleading?

["Even the normally very helpful National Center for Education Statistics (NCES) of the U.S. Department of Education has done little to improve the quality of statistics on high school completion." (Greene, 2002, p.1)]

["The results are consistent with high school completion rates reported by the NCES...but this report expands upon the NCES report by providing graduation rates for states, districts, and ethnic/racial sub-groups that are not provided by the NCES." (Greene, 2002,p.9)]

The primary role of NCES is to collect data. Indeed, it has long been argued by some that NCES should do nothing more than collect data, organize them, and make them publicly available; any expenditure of NCES funds for any other purpose only detracts from their primary mission. To be fair, the overwhelming majority of NCES's efforts have always been devoted to that primary mission. But, the agency has also tried to make those data more accessible and understandable to the public by publishing compendia and descriptive reports filled with tables, charts, and graphs. The total volume of these publications over the years would easily fill several warehouses.

Nonetheless, NCES does not put every possible combination of measures into a table, chart, or graph every year. The annual *Digest of Education Statistics* is a heavy book, filled with hundreds of the most basic tables. If it were to attempt to cross tabulate all the measures available, it would be billions of pages thick.

To accuse NCES of "not providing" data because they have not printed every possible combination of measures in published crosstabulations is rather unfair. Anyone can, with less than a half-hour's worth of effort, produce a table of state-by-state completion ratios, using graduation numbers by state available from NCES for the numerator, and age cohort numbers available from NCES or the U.S. Census Bureau for the denominator.

Moreover, it is simply not true that NCES never publishes this type of table. Anyone conducting a thorough search would probably find plenty. Here's a link to one I just happen to have handy because I made it:

http://nces.ed.gov/pubs/esn/n23cb.asp

For another example, a casual look back through the years of the OECD's annual *Education at a Glance* statistical compendium will reveal that they have been publishing completion ratios for over a decade.

Even completion ratios, despite their superiority in some respects, are not perfect measures, nor would they be completely reliable to use in judging the performance of any high school, aside from the one on the deserted desert island. There remains the problem of student migration, and the unfairness of judging one high school for the progress of students who transfer there sometime in between the start and the finish of their high school careers.

Probably the most valid and reliable measure one could use for Greene's expressed purposes with the current batch of statistics available to at least some U.S. jurisdictions is a high school completion ratio with each student assigned to different schools weighted by the amount of time spent in each school. That would be much fairer than Greene's statistic, but is still not possible to calculate in most of the United States, where the requisite improvements in accounting for student migration remain to be implemented.

The Census Bureau's work is suspect, too?

["Other factors may explain the modest differences between my graduation rates and the NCES high school completion rates.... CPS relies upon self-reported educational status for NCES to compute high school completion rates.

That is, people have to describe honestly to the survey researchers whether they received a high school diploma." (Greene, 2002, p.7)]

Ah, so the Census Bureau survey respondents might lie, presumably from embarrassment, to the contract employee at the call center, even after being assured that the law prohibits any personal identifying information from being revealed and after agreeing to participate and tell the truth. If Greene's cautions regarding these survey data were valid, we should reasonably question most of the data in most Census Bureau collections.

In fact, however, federal statistical agencies conduct a steady stream of studies in which they investigate the reliability of responses under a variety of conditions.

No matter, if Greene would prefer to trust administrative records, he can use graduation numbers reported by school districts as the numerator in a completion ratio. Indeed, that is the figure normally used. Then, he can divide by the number of persons in an appropriate age cohort, unless he thinks survey respondents are prone to lie about their age, too.

Implausible dropout statistics

["This report also improves upon state and district reported dropout rates, which unfortunately often implausibly understate problems." (Greene, 2002, p.9)]

Dropout statistics reported by school districts to their state education agencies tend to understate the real number of dropouts. But, they are not "implausible." Indeed, they are exactly and nothing more than "plausible." Reported dropout numbers tend to be the lowest possible *plausible* number schools and districts can get away with providing. Where schools and districts are only required by regulation to report the number of students who signed up in early September as their enrollment, and the only subtractions from the rolls they may feel compelled to report are for those students who transfer to another school within the same district (where they are claimed on the rolls there), those are likely to be the only subtractions reported. Where schools and districts are required to report as dropouts those students they cannot account for in between September enrollment and later "average daily attendance" counts, those are likely to be the only ones they report.

Implausible inclusion of GED graduates

["Recipients of GEDs are not, properly speaking, "graduates" of any high school....cannot be credited to the high school. Similarly, a doctor cannot claim as "cures" patients who have transferred to other doctors for treatment." (Greene, 2002, p.6)]

The Manhattan Institute also jumps on the GED-bashing bandwagon, arguing that those who drop out of regular high school, then pass the General Educational Development (GED) Exam (a.k.a., high-school equivalency exam), should not be counted as high school graduates. Why? Two reasons, employers do not seem to give the GED much credence (which, by itself, does not invalidate the worth of the GED program itself but, rather, its lousy image, which the Manhattan Institute does its best to perpetuate), and two, GED exam passage should not be credited to the high school left behind.

What if a student was doing well in high school, but leaves in the last semester due to some major event (e.g., death of family breadwinner, sickness, pregnancy) and later passes the GED. That's hardly an exceptional circumstance. The high school left behind should get no credit whatsoever for that student's progress? Instead, it should be blamed for that student "dropping out?" That hardly seems fair.

What to do?

There are so many systematic biases at play in what the Manhattan Institute cobbled together that the statistic should not be considered legitimate. Their "graduation rates" are neither valid nor reliable for their suggested use. If they were to actually be used to make judgments, two types of jurisdictions would be the most likely to be unfairly judged:

- those with rising fertility rates will see an overcompensation for their migration patterns because disproportionally rising enrollments in the primary grades will unfairly increase 8th grade enrollments after adjustment, and that will artificially lower graduation rates; and

- those with transient populations, school choice, declining populations overall, and states with relatively small and/or supervisory districts (and, thus, more transfers) will have 8th grade enrollments that are too high and not adjusted downward enough for the "temporal decay" of enrollment's reliability, resulting in artificially lower graduation rates.

No one can know if these two different biases (caused by using total student population changes to adjust 8th grade enrollment and by relying on enrollment, rather than attendance numbers as the rate base) would cancel each other out, as Greene claims they do, without some empirical simulations. It is difficult to imagine how they could, though. Each of the biases would be most detrimental in very different regions of the country and each would seem to pull in the same, not opposite, direction for the most part (and underestimate the true graduation rates). The first bias would likely be most prominent in the suburban Sunbelt and some poor immigrant communities. The second bias would likely be most prominent in highly transient communities and declining urban centers of the Rustbelt. But, there exist also some regions where the biases are likely to be compounded (think Miami or Phoenix). The method is likely to lead to underestimated graduation rates nationwide at the front end of rises in fertility rates, like the baby boom or baby boom echo, and overestimated graduation rates at the tail ends.

Begging the reader's indulgence let me provide just one simple example with some numbers in it. Consider the fictional school district Sunbelturbia. With new housing developments just completed, and in influx of new young families, the district now has both a positive population growth rate and a positive fertility rate. Table 1 lists Sunbelturbia's enrollments by grade (K through 12) in a base year (year 0) and four years later. For the moment, assume that no one drops out of school in Sunbelturbia.

Table 1: First hypothetical enrollment data set

Grade Level	year 0 enrollment	year 4 enrollment
K	10,000	14,000
1	10,000	13,750
2	10,000	13,500
3	10,000	13,250
4	10,000	13,000
5	10,000	12,750
6	10,000	12,500
7	10,000	12,250
8	10,000	12,000
9	10,000	11,750
10	10,000	11,500
11	10,000	11,250
12	10,000	11,000
Total	130,000	162,500

Employing the Greene Method, first we calculate the percentage change in Sunbelturbia's total enrollment in the four years (that comes to 25 percent, or .25). We use that result to adjust the year 0 8th grade enrollment (10,000 + (10,000 * .25) = 12,500. Now, we divide the year 4 12th grade enrollment by the year 0 8th grade adjusted enrollment (11,000 / 12,500) = .88 (i.e., a graduation ratio of 88 percent).

The Greene Method estimates that only 88 percent of year 0's eighth graders graduate four years later in Sunbelturbia. It estimates that 12 percent, or 1,200, of Sunbelturbia's year 0 8th graders have dropped out by 12th grade, even though none have. Indeed, with the numbers all there, we can see that 1,000 more students graduated in year 4 than were in 8th grade in year 0. The true graduation ratio in Sunbelturbia is 110 percent.

But, that does not account for student migration. Still assuming no dropouts, the 1,000 extra students in year 4's 12th grade class are all in-migrants to Sunbelturbia, having arrived in the school district sometime between year 0 and year 4. Adjusting for the in-migration, the true graduation

ratio in Sunbelturbia is 100 percent (remember, no dropouts).

The "quite accurate," "reliable and straightforward" MI estimate is off by 22 percentage points, not accounting for migration, and by 12 percentage points accounting for migration. Moreover, it does not even move in the right direction, estimating an enrollment decline for a cohort that is either increasing in size or stable in size.

Now, let's give the fictional Sunbelturbia a dropout problem. Assume compulsory school requirements end at grade 11. We know from table 1 that, by 12th grade, the year 0 8th grade cohort has increased in size by 10 percent due to in-migration. In table 2, that proportion remains the same, but with the dropout problem added to the mix.

Table 2: Second hypothetical enrollment data set

Grade Level	year 0 enrollment	year 4 enrollment
K	10,000	14,000
1	10,000	13,750
2	10,000	13,500
3	10,000	13,250
4	10,000	13,000
5	10,000	12,750
6	10,000	12,500
7	10,000	12,250
8	10,000	12,000
9	10,000	11,750
10	10,000	11,500
11	7,500	8,625
12	5,000	5,750
Total	122,500	153,875

In a high-dropout Sunbelturbia using the Greene Method, the percentage change in total enrollment is 25.6 percent, the adjusted year 0 8th grade enrollment is 12,560, and the calculated graduation ratio is 45.8 percent.

But, we have all the "actual" numbers and so can calculate the true graduation ratio high-dropout Sunbelturbia. Assuming that in-migrants and non-migrants drop out at the same rate, there would be 5,227 members of the year 0 8th grade cohort left at graduation time. The "actual" graduation ratio is 52.3 percent.

The MI estimate is off by 6.5 percentage points. Again, it is not even close.

Comparing MI ratios to traditional completion ratios

The MI claims that their "graduation rates" are similar to the completion ratios most statisticians would use for the same jurisdictions, which only begs the question: why not just use the latter? ...and throw away the MI convolution? There are at least four general threats to the validity and reliability of the MI's ratios:

- changes in the demographic trend over age cohorts;
- the "temporal decay" in enrollments during the school year;
- inconsistent reporting dates, standards, requirements, incentives, efforts, and so on across jurisdictions; and
- student migration.

With simple completion ratios–number of grads in a year divided by the size of that age cohort–the only threat to the validity of the statistic is student migration. A statistic with one very clear, definable–and fixable–problem is far better than a statistic with at least four major problems, probably a bunch of minor ones, and no hope of being fixed anytime soon.

The simple completion ratios most statisticians use divide a number of graduates by the number of persons of their same age living in the same jurisdiction. In effect, simple completion ratios take dropouts directly into account–they are the difference between the number of graduates and the total number of persons in the age cohort.

The MI method does not account for dropouts. Nor does it account for fertility rates. Nor does it

account for migration, except in a very crude, indirect way.

Moreover, the statistical biases mentioned above are just the tip of an iceberg's worth of problems. MI also breaks student populations into ethnic subgroups and makes the same reassuring claims regarding the reliability and validity of his graduation rates for them. Adopt the MI method, however, and statistical "paradoxes" will likely pop in profusion. One will witness Simpson's Paradox (caused by comparison of changes in subgroups with unstandardized bases) and Kelley's Paradox (a variant of regression toward the mean), and probably others.

It is difficult to estimate how much damage and confusion widespread adoption of the MI Method would cause, but it likely would be substantial.

Pioneering work

["The lack of candor...is a fundamental problem in education." (Greene, 2002, p.9)]

["The relative inattention devoted to graduation rates is at least partly explained by the confusing, inconsistent, and sometimes misleading way in which the rate of high school completion is measured." (Greene, 2002, p.1)]

The MI report is praised by the author of its preface as "pioneering." Strictly speaking, the preface writer is correct, the report is "pioneering." But, there's a reason for that.

In fact, education statisticians have discussed these methodological issues (that Greene seems to think he has freshly uncovered) among themselves *ad infinitum* for decades. Greene's method is, as he admits, simple and, if it were also a good method, the folks at NCES would have adopted it long ago. They did not adopt it because it is not a good method; the far more reliable completion ratio is simpler, easier to compute, and easier to understand. But, even completion ratios are not valid for making the type of judgments the MI favors if they do not account for student migration.

Greene can chastise the National Center for Education Statistics all he wants. But, they are not at fault for the problems inherent in calculating valid and reliable graduation or dropout rates; they only collect the data, they do not produce them. Nor do they have the direct power to fix the problem. Nor are they responsible for the U.S. Constitution, which cedes all education authority to the states, to each to manage on its own and in its own way.

The only way to fix the problem requires a far greater invasion of privacy than many U.S. politicians are willing to suggest imposing. The Manhattan Institute is welcome to propose a federal law that would impose uniform statistical reporting standards across the states, and the type of external monitoring and tracking of individual students' comings and goings that would fix the methodological problems, as others have before them. Then, they likely will be labeled "Big Brother," too.

Some state statistical agencies have been making the effort to convert their databases so that students, not schools or school districts, become the lowest unit of analysis. They hope to be able to track each and every student as they either stay put or migrate from grade to grade and school to school. Such a monitoring and tracking system is necessary for calculating valid and reliable graduation rates of the sort Greene, and many others, hope for. It remains to be seen, however, how many U.S. states will attempt such a system; probably less than a dozen have thus far. It also remains to be seen how well these systems will track students in their teenage years, when they become more willful and recalcitrant, but the schools still desire the state aid that depends upon their being present, at least in statistical form. Finally, it also remains to be seen how similar and comparable these systems will be across states if they do become more popular.

States that have successfully completed the transition from student headcounts based on school district submissions of aggregate figures to school district submissions of individual student records have witnessed dramatic jumps in their statistical time series. Headcounts based on microdata have,

in some states, differed substantially from earlier headcounts based on macrodata. These discontinuities probably elicited little surprise from the experienced data warriors at NCES but should rattle confidence at the Manhattan Institute which has assumed more stability to these numbers than is warranted. Indeed, some of the state conversions from macro to microdata occurred within the time frame of the Institute's calculations.

A better alternative to the Manhattan Institute report would have been an essay that laid out the rather compelling arguments for why completion ratios are, for many purposes, superior measures to some so-called and self-reported graduation rates. Alternatively, the Institute could have encouraged all states to convert to pupil-centered data systems.

With these efforts, many of the fastidious and hard-working statisticians at NCES and the Census Bureau would likely have concurred. Their utterances would have been expressed in private, of course, because, as public servants, they are prohibited from speaking out, defending themselves, or otherwise behaving like advocates. That prohibition, of course, makes them easy targets for those prone to pick on them.

CONCLUSION & DISCUSSION

Graduation rates and completion ratios are fundamentally important measures of education system performance. Indeed, they can serve as a legitimate No Child Left Behind Act supplemental indicators of school success. All the more reason that those most knowledgeable should be the ones to interpret them to policy makers and the public.

The fact that these statistics are so problematic is not the fault of the education statisticians who calculate them, as the Manhattan Institute suggests. Indeed, our country is privileged to have in residence many genuine experts on this topic–statisticians with decades of practical experience working with these data and improving them–smart people who know this stuff cold. But, they do not work in think tanks and they do not cater to the media as think tanks do. They just do their jobs, and their work is routinely ignored.

The Manhattan Institute embarked on this work when they noticed that graduation rates across U.S. states and districts seemed grossly inflated. It is apparent from their report that they simply did not understand the difference between graduation rates and completion ratios. In other words, they lacked the most rudimentary understanding of how these statistics are defined and how they are collected. But instead of inquiring of the genuine experts in order to gain an understanding, they made assumptions, jumped to conclusions, and went quickly to the microphones.

Most of the blame and shame that fills the MI report emanates from their own confusion of the terminology and collection processes. All of the blame and shame that fills the report could have been avoided had the MI simply taken the time to make inquiries and attempt to understand a world unfamiliar to them.

Unknown to the Manhattan Institute, some of our country's most knowledgeable and experienced education statisticians have been laboring for years to adopt and implement pupil-centered data systems. It has been an epic struggle against bureaucratic inertia, short-sidedness, and the typical under-appreciation of the value of public statistics.

Nonetheless, these education statisticians have made substantial progress. By the year 2002, when the MI report was released, over a dozen states had successfully completed the arduous conversion to new pupil-centered data systems.

Those genuine statistical experts should be congratulated. Instead, they received accusatory telephone calls from reporters convinced by the MI report that they were either dishonest or incompetent. Instead of being treated deservedly, as heroes, they were vilified unfairly as near-criminals. Instead of being thanked for their hard work and dedication to improving public information, they were kicked in the teeth.

Meanwhile, the Manhattan Institute congratulates itself for discovering a problem that more knowledgeable folk have been aware of for decades, and congratulates itself for helping to fix a

problem that more dedicated folk have been working on for many years now. The Manhattan Institute surely will entreat potential benefactors to donate more funds so that they may continue to pursue more of their noble efforts to make the world a better place.

Appendix A: Rates versus Ratios

An illustrative example of the gulf in understanding between full-time education statisticians and others is illustrated by the relatively arcane distinction between rates and ratios. It can cause what passes for heated debate among education and population statisticians that is almost certain to bore most anyone else. So, the distinction is widely ignored. The statistical purist's requirements for calling a measure a "rate" are more stringent than those for calling a measure a "ratio." A *ratio* need be nothing more than two numbers arranged in the form of a division algorithm–a dividend over a divisor. A population *rate* requires that all the particular individuals counted in the numerator also be present in the denominator count.

Rates are often considered purer representations of population dynamics and, in theory, they are. In practice, however, carefully crafted ratios are often far more trustworthy than the best approximation to a rate that can reasonably be computed with extant data. So it tends to be most often with graduation rates and ratios.

Think of the requirements for a reliable graduation rate. An appropriate numerator might be easy enough to obtain–the number of graduates reported in administrative records, for example. But, what would one use for the denominator? The number of students enrolled the final semester of high school? That would simply provide a measure of the success of students that particular semester, and most citizens might like a measure that represents a longer term process. For example, a denominator appropriate for holding a high school responsible for its success in getting its students through might be the number of students who enroll at the beginning of the first year at that high school.

Even that could only be a pure measure of "success" if the high school were the only one on a deserted island. In the rest of the world, students can move around, and no where more so than in the highly mobile United States of America. Much moving around occurs in the first weeks of school, when some parents and students shop around. A student may attend the first week of classes in one school, hate it, and then enroll somewhere else. But, some moving around occurs during the year, too, and much in between school years. Families move. Some students drop out. Some students transfer to different schools.

The incidence of moving around is not consistent across states or school districts. There are clear, systematic biases. There's more moving around among poor families and other families who rent, rather than own, their homes. There's more moving around among high-level corporate professional, military and diplomatic families. There's more moving around in states and districts with more school choice, be that due to the availability of charter, religious, or private schools, or a public school open enrollment scheme. There's more moving around across school districts in states with smaller districts. There's more moving around in urban areas where several schools are close by, than in isolated small towns where only one is. There's more moving around in poorer neighborhoods where students might take time off to work and support their families, and then, sometimes, return to school later. There's more moving around in jurisdictions with recent closures of public housing projects, or recent openings of large, new housing developments. To hold a particular high school district responsible for the graduation of a student who only transfers there during her senior year seems rather unfair.

Not only are there systematic biases to the incidence of student migration and persistence across districts and states, there are systematic biases across time. Migration rates change with the ups and downs of local, regional, and national economic and housing conditions, with changes in governmental social program policies, and with changes in U.S. military commitments.

Moreover, if ANY of the students who move into a jurisdiction after the first week of the first year are counted in the graduation rate calculation described above, the resulting statistic cannot validly be called a rate, at least not by the standard of the most reputable statistical agencies. Remember, in a valid graduation *rate*, every graduate counted in the numerator must also be counted in the denominator. In the case of transfer students who arrive after the point in time when the denominator is set, that cannot be the case.

Some districts and states currently attempt to make some adjustment for student migration in their school- and district-level accountability schemes. One method is to weight each student's progress in a grade at a particular school by the proportion of the year spent at that school. This doesn't account for the debilitating effects of the transfer itself, but it is a fairer method than the usual alternatives, which assume that students never move. The jurisdictions with data detailed down to the student level, however, may not have the legal clearance to provide it to outsiders. Americans, more so than their counterparts overseas, can be pretty fussy about the dissemination of individual-level information.

Appendix B: A Brief Primer on Student Headcounts

What are enrollments, exactly? Traditionally, they are the names of students who *sign up* at the beginning of the school year with the expressed intention of attending a particular school–the names on the roll. Any student is enrolled whose name is listed on the sign-up sheet. Not all students who sign up, however, show up. Not all students who sign up and show up, stay.

In fact, what NCES labels "enrollment" in its more popular consumer publications, such as the Digest of Education Statistics, is actually what most states themselves label "membership." Unlike pure enrollment, membership attempts to account for student placements outside home school districts. Say, a student desires or needs an educational program available in a neighboring school district, but not her own. She then enrolls in the neighboring district, which is reimbursed by the home district for her cost, but she is included in the home district's membership count. At least that is how it is supposed to work. Some states count membership this way, and some do not.

So, states vary in the character of their enrollment (i.e., membership) collections somewhat, but most pinpoint some date in September or early October when districts are supposed to count the number of students who are listed in their enrollment log at that point in time. Districts vary quite a lot, however, in the level of effort they expend toward de-listing students who never showed up for classes, or who quit coming after the first week of school

Greene implies that districts have an incentive to fudge the dropout and graduation numbers. But, most of them do not have much incentive to keep the enrollment numbers accurate or up-to-date, either. In states where state aid allocations to districts are made based on their enrollment counts, the districts face a severe *disincentive* to keep their enrollment numbers up to date (except in regard to the students transferring *into* their district during the first few weeks of classes). The end result is that aggregate enrollment figures overestimate the true number of students, and even double count some students in the aggregate.

In some jurisdictions, however, a fuss is made over "attendance" counts, which vary widely (some would say wildly) across states in their definitions, reporting standards, and date of collection. Attendance is a count of the number of students who *show up* on a given day. Some districts have been known to make these days--when designated officials count heads--into something like school-based holidays, with good food, movies, and games, pulling out all the stops to get enrolled students, even those who have already dropped out of school, to show up. Attendance numbers are more important in states that base their state aid allocation on attendance counts.

As one might suspect, some states are more vigorous than others in their effort to attain valid and accurate enrollment or attendance numbers. States with more power concentrated centrally, smaller geographic size, or a larger responsibility for school funding are likely to be more able and more diligent in checking the accuracy of the student headcounts supplied by their districts.

If both enrollment and attendance counts were reliable, one should expect to find a strong correlation between the two numbers across states, and one does. But correlation statistics tell one about parallel movements of entire masses of data. At the margins, these statistics do not seem so well matched.

Using 1999 state-level data, I subtracted attendance counts from enrollment counts and observed those differences across states. The mean percentage difference (using attendance as the base) was eight, with a range of over fifteen percentage points. For the reasons mentioned above, one would expect that enrollment would always be higher than attendance (it is not), and that the differences would be higher for states with declining populations and lower for states with increasing populations. That is because enrollment is almost always collected earlier in the year than is attendance.

For the extreme cases, the pattern seems to hold. The District of Columbia, with a generally declining population had the largest difference between enrollment and attendance counts (18 percent) and fast-growing New Mexico and Virginia actually showed headcount increases between their enrollment and attendance counts. But, a further look finds the pattern breaking down completely. Arizona, Nevada, and Oregon, three of fastest growing states in 1999 are among the states with the largest differences. Iowa, North Dakota, and West Virginia, all slow-growing states, are among those with the smallest differences.

There is a jumble of divergent factors making these counts behave badly, including large variations in definitions, reporting requirements, effort, and transparency across states, and demographic trends. This volatility of total student population statistics across states from one year, could only be exacerbated at smaller aggregations, such as the more relevant grades eight through twelve, or at the school district level, and when accumulated over four years, as happens using the Manhattan Institute method.

REFERENCES

CBS News (2004). The 'Texas Miracle', *60 Minutes II*, August 25.

CNN Presents (2005), Jay Greene defends NCLB and explains dropout rates, May 14.

Greene, J. J. (2002a). *High School Graduation Rates in the United States*, Manhattan Institute, Center for Civic Innovation, November 2001 (Revised, April 2002). Available online: http://www.manhattan-institute.org/pdf/cr_baeo.pdf

Greene, J.J. (2002b). *High School Graduation Rates in Washington State*, Manhattan Institute, Center for Civic Innovation, August, 2002.

Greene, J.J. & Hall, J.C. (2002c). *High School Graduation Rates in Ohio*, Buckeye Institute.

Greene, J.J. & Forster, G. (2003). *Public High School Graduation and College Readiness Rates in the United States*, Manhattan Institute, Education Working Paper #3, September.

National Public Radio (2005), Governors Move to Standardize Graduation Figures, *All Things Considered*, July 19. Available online: http://www.npr.org/templates/story/story.php?storyId=4759634

Organisation for Economic Co-Operation and Development (any year) *Education at a Glance: OECD Indicators*. Paris: OECD.

Organisation for Economic Co-Operation and Development (1999). *Investing in Education: Analysis of the 1999 World Education Indicators*. Paris: OECD.

Quality Counts (2002) Education Week, Editorial Projects in Education, Bethesda, MD.

Notes

Thanks are due the experienced and wise Karla Carr, John Sietsema, Bruce Thompson, Larry Rudner, and four anonymous referees for their editorial suggestions and clarifications of some issues regarding the statistics described herein, and the idiosyncracies inherent in their collection. None of them, of course, is the least bit responsible for what I write.

Citation

Phelps, Richard P. (2005). A Review of *High School Graduation Rates in the United States*. *Practical Assessment Research & Evaluation*, 10(15). Available online: http://pareonline.net/getvn.asp?v=10&n=15

Author

Richard P. Phelps grew up in St. Louis, Missouri and received degrees from Washington, Indiana, and Harvard Universities, and a PhD from Penn's Wharton School. He taught secondary school mathematics in Burkina Faso (West Africa); worked as an education statistician at the Organisation for Economic Co-operation and Development in Paris, the U.S. General Accounting Office, Westat, and Indiana's Education Department; and has published two books, several statistical compendia, and dozens of articles in scholarly journals. Phelps is a co-founder of the Nonpartisan Education Review.