Computation Skills, Calculators, and Achievement Gaps: An Analysis of NAEP Items

Tom Loveless
Brown Center on Education Policy
The Brookings Institution
April 2004

DRAFT: DO NOT QUOTE OR CITE WITHOUT AUTHOR'S PERMISSION

Paper prepared for the 2004 American Educational Research Association Conference, San Diego, CA, April 12-16, 2004.

I am grateful to Alice Henriques and John Coughlan for research assistance. I would also like to thank the National Assessment Governing Board and Educational Testing Service for providing the data used in this paper.

This paper uses data from the National Assessment of Educational Progress (NAEP) to study three issues in K-12 mathematics. It examines national trends in computation skills, investigates whether allowing calculators on NAEP items produces significantly different results compared to not allowing calculators, and analyzes the impact of allowing calculators on the performance gaps among black, white, and Hispanic students.

All three topics are controversial. Computing quickly and accurately is recognized by most math educators as an essential skill, but the question of how much emphasis computation should receive in the K-12 curriculum provokes a heated debate. In its 1989 standards, the National Council of Teachers of Mathematics (NCTM) recommended de-emphasizing computation skills, warning "clearly, paper and pencil computations cannot continue to dominate the curriculum." The NCTM standards had a profound impact on state curricular policy. By 2000, all but two states, California and Massachusetts, modeled their own curriculum standards on the NCTM's, and publishers revised math textbooks to conform with NCTM's prescriptions.[1]

The call to de-emphasize basic skills drew intense criticism from professional mathematicians and parent groups. They argued that the algorithms of arithmetic prepare students for more sophisticated topics and that students apprehend the fundamental structure of mathematics through the mastery of such seemingly trivial operations as long division.[2] In 2000, the NCTM released a revised set of standards, *Principles and Standards for School Mathematics*. In deference to critics, the language on computation was softened. By the end of fifth grade, the document declared, "students should be computing fluently with whole numbers." [3] Notwithstanding the document's more moderate tone, the NCTM standards did not abandon the position that computing skills should be de-emphasized.

The national policy influence of the NCTM is reflected in the mathematics framework of the National Assessment of Educational Progress (NAEP), also known as "the nation's report card."[4] The program periodically administers two tests—the main NAEP, given since 1990 in mathematics, and the long term trend (LTT NAEP), given since 1973. The main NAEP is governed by a framework that largely mirrors the NCTM's views on the K-12 mathematics curriculum.[5] The main NAEP test, for example, allows students to use calculators on a portion of the test, includes extended-response items which ask students to explain their mathematical reasoning, contains items on which students may receive partial credit, and provides manipulative materials for students to use on selected items. In addition to an overall score at the fourth, eighth, and twelfth grades, main NAEP scores are reported on NCTM's five strands—number and operations, measurement, data analysis, algebra, and geometry. Consistent with the notion of de-emphasizing paper and pencil skills, no score is reported on arithmetic or the ability to compute, even at fourth grade.

In contrast to the main NAEP, the long term trend NAEP remains essentially unaltered since its inception in 1973. Last given in 1999, the test reflects a "pre-standards" view of mathematics. All items are multiple choice. A significant proportion of items are devoted to arithmetic and computation skills, but the LTT NAEP does not report disaggregated scores by skill area. In other words, the public is in the dark on whether today's students know how to compute accurately. The main NAEP and most state assessments do not assess students' computation skills. The study at hand will examine selected data from the LTT NAEP to shed light on American students' computation skills.

The subject of calculators in the classroom is another flashpoint in debates over the modern math curriculum. Most observers agree that technology has the capacity to produce exciting new tools for learning mathematics. Nevertheless, and despite being commonly used in contemporary classrooms, calculators are a never-ending source of disagreement. Arguments rage over how and when calculators should be used in instruction, especially with young students. Testing officials wrestle with whether to allow calculators on state assessments.[6]

Again, the NCTM plays an important role in the controversy. The NCTM first expressed its support for calculators in 1974. It reissued the endorsement in 1980, calling for schools to "introduce calculators and computers into the classroom at the earliest grade practicable." In their 1989 standards , the NCTM recommended that calculators be used in grades K-4. Powerful allies rallied to the NCTM's side. In 1990, the National Research Council issued "Reshaping School Mathematics," a report that urged "the replacement of most paper-and-pencil drills with calculator-based instruction" starting in kindergarten. Because calculators "diminish the role of routine computations," the report advised, "young children can instead be given activities with calculators that emphasize discovery and exploration." Since the early 1990s, the National Science Foundation has awarded tens of millions of dollars for the development of curriculum materials that promise to integrate calculators into instruction.[7]

Critics believe calculators may impede learning, especially when used by students who haven't memorized basic facts (e.g., 2+2, 6x7, 14-9) or learned how to add, subtract, multiply, and divide on paper. The risk is that calculators will become a crutch for students. Worse yet, children may never acquire a deep understanding of how numbers work if, on first exposure to mathematical operations, they merely push buttons to arrive at answers.[8] The politics of the issue change as the focus shifts from education schools to the frontlines of schooling. Surveys show that professors in schools of education believe calculators should be used more often in teaching math. But teachers want them used less, and a large majority of the public thinks that they shouldn't be used at all with young children.[9]

An extensive literature on calculators exists. More than 120 studies compare the test score changes of treatment groups who use calculators with control groups who do not use them, with comparisons in achievement made after both groups have studied the same math curriculum. Three meta-analyses of the research give qualified support for using calculators in instruction (Hembree and Dessart, 1985; Smith, 1997; and Ellington, 2003). They generally agree that calculators do not negatively affect paper-and-pencil skills and have a positive effect on students' attitudes toward math. However, in the Ellington meta-analysis fewer than 20% of the studies featured students in the fourth grade and lower, the very years in which basic arithmetic is taught. Ellington's advice on using calculators is tempered accordingly; "Because limited research has been conducted featuring the early grades, calculators should be restricted to experimentation and concept development activities."[10]

A design question in these studies—and a significant issue for NAEP—involves selecting an appropriate instrument for measuring the effects of calculators in experiments. If calculators are an integral part of instruction with treatment groups, then measuring learning gains with an assessment that allows for at least some calculator use seems reasonable. Not allowing them could bias findings against calculators. If, on the other hand, calculators are allowed on such assessments, any advantage detected from treatment over control groups may in fact be due to increased facility using calculators, not gains in computation skills. Findings could be biased in favor of calculators.

What do the different testing protocols show? Thirteen studies have evaluated computation skills with calculators allowed on the assessment used to measure gains. The mean effect size is positive and significant (.43, $p<.01$). Fifteen studies have evaluated computation skills without calculators allowed on the assessment. The mean effect size is statistically insignificant (.03). Allowing calculators on tests appears to influence the estimates of calculators' effects on students' computation skills. How does this relate to NAEP? The main NAEP allows calculators on 35% to 40% of items on the fourth grade test. Most states allow them on tests linked to accountability systems. Calculators may produce misleading information on students' ability to compute. The extent to which calculators influence the assessment of computation skills on NAEP items is investigated below.

Gaps in math achievement among racial and ethnic groups also figure prominently in analyses of NAEP scores. On the long term trend NAEP, the difference in black-white test scores narrowed significantly from 1973 to the late 1980s and then began widening again in the 1990s. Lee (2002) points out the major trends in the gap:

> During the period between 1971 and 1986/1988 when the achievement gap between Whites and Blacks narrowed, White students' achievement level was quite flat, whereas Black students made substantial academic gains. In

> contrast, during the period between 1986/1988 and 1999, when the gap grew, the pattern reversed: White students improved their achievement but Black students made few gains on the NAEP. Consequently, the narrowing of the Black-White achievement gap stopped and in some cases the gaps returned to the level of the late 1970s or early 1980s. (Lee, 2002, pp. 3-4) [11]

The Hispanic-white gap evidences no consistent pattern. Like blacks and whites, Hispanics made gains on the NAEP from 1973 to 1999. But their scores were more erratic, with scores moving in the same direction—whether up or down--on very few adjacent assessment years.[12] It must be kept in mind that the relatively smaller sample sizes of both black and Hispanic students increases the chances of score volatility on NAEP.

Research Questions, Methods, and Limitations

This study updates and extends a study of computation scores conducted by the Brown Center on Education Policy in 2002.[13] The previous study did not examine racial and ethnic differences in student performance. The current study examines these performance gaps. The previous study was based on data that did not separate items on which calculators were allowed and not allowed. The current study makes separate analyses of these two test conditions. The current study also analyzes how nine year olds performed on a subset of matched items, examining how well a group of students who were allowed to use calculators were able to compute compared with students who were not allowed calculators on the same items.

The study investigates six research questions (the analytical strategy appears in parentheses). The first four are:

1. Did computation skills improve, decline, or remain the same from 1982 to 1999 (analysis of gain scores)?
2. How do the 1980s and 1990s compare on trends in computation skills? This analysis compares scores in the decades immediately before and after the release of the NCTM standards.
3. How sound were students' computational skills as of 1999, the last time the trend NAEP was administered (analysis of absolute level)?
4. In 1999, how large were black-white and Hispanic-white gaps in computation skills? Have they changed over time (analysis of absolute level and gain scores)?

Two additional research questions are investigated using a subset of matched, calculator and non-calculator items. On these items, a randomly selected subsample of NAEP nine year olds were provided calculators and allowed to use them to compute answers. (Note that these items do not count in the calculation of the national NAEP score).

5. Given an identical set of computation items, how do students perform when they are allowed to use a calculator compared to when a calculator is not allowed?
6. Do the conclusions of the analyses above (comparison of the 1980s and 1990s, progress in closing black-white and Hispanic gaps) differ when calculators are available or not?

Computation items from the NAEP were initially identified based on short descriptors provided for each item. Items were also examined with the assistance of staff of the National Assessment Governing Board (NAGB), the organization governing NAEP, at the Washington office of NAGB. To qualify for the study, items had to ask students to compute using one of four operations (addition, subtraction, multiplication, and division) with four types of number (whole numbers, fractions, decimals and percentages). Word problems were excluded. Eleven items met the criteria on the test for nine year olds, nineteen items on the test for thirteen year olds, and twelve items on the test for seventeen year olds. The data for student performance were provided by NAGB and the Educational Testing Service (ETS).

There is an important limitation to this type of study. The analysis is based on a subset of items from a larger test and therefore may not adequately sample the domain of a particular skill. This is less a problem for drawing conclusions about computation skills as a whole—on which several items exist across three or more skill areas--but it is a fault to which individual skill clusters are vulnerable. A good example is that only two division computation items were identified on the test for nine year olds. They obviously were not included on NAEP with the intention of comprehensively measuring nine year olds' ability to divide accurately with whole numbers. They are part of a larger test designed to sample, not to exhaustively inventory, children's basic mathematics skills.

That said, these NAEP data possess several strengths. No national test of computation skills exists, so even if the findings are limited to student performance on these items alone, the data are valuable in offering a limited portrait of national performance. The main NAEP has very few computation items, so the LTT NAEP provides the best evidence available on the topic. The LTT NAEP has been given to large, random samples of students for more than three decades, producing comparable scores that allow for analyses of trends. And, as already noted, a matched set of items on the NAEP LTT allows for a comparison of student performance with and without the availability of calculators.

A final word on methods. Readers will notice that standard tests of statistical significance were not performed on changes in student performance. This is because the analysis is based on clusters of test items. Standard errors were only available for the individual items and not for clusters (i.e., for a two item cluster, calculating the standard error would require the probabilities of getting both items correct, one or the other item correct, and neither item correct). I compensated for this by running significance tests on individual items and checking to see if they confirmed the basic story told by the clusters comprising those items. They did. The significance tests for the matched-item calculator analysis are provided in Appendix A.

Trends in Computation Skills

Table 1 presents NAEP data on student computation skills from 1982 to 1999. Items are clustered by skill. The "scores" reported for each skill cluster represents the mean percentage of students answering that cluster's items correctly in 1982, 1990, and 1999. In a skill cluster comprising only two items, for example, if 30% answered one item correctly and 70% answered the other one correctly, the reported score would be 50%. The final three columns show changes in scores for different time periods. The first two columns disaggregate changes by decade (actually, changes from 1982 to 1990 and 1990 to 1999). The final column shows changes over the entire period of the study, from 1982 to 1999. All of the data in Table 1 are based on student performance without the aid of calculators.

Insert Tables 1 and 1A

As the last column in Table 1 indicates, computation skills generally improved from 1982 to 1999. The greatest gains were in seventeen and thirteen year olds' ability to compute with percentages (gains, respectively, of 13.7 and 7.6 percentage points). However, thirteen and seventeen year olds both declined in the other two areas of computation. For thirteen year olds, the ability to compute with whole numbers (-3.2) and fractions (-5.4) fell. For seventeen year olds, the declines for computation with fractions was -11.4 and with decimals was -5.0. Nine year olds gained in every area of computation examined in the current analysis.

Table 1A presents a scorecard for the data in Table 1. It indicates that the 1980s were a better decade for computation skills than the 1990s. Of trends in the ten skill clusters, nine favor the 1980s and only one favors the 1990s. Table 1 reports that in the 1990s computational proficiency declined across all ages and skills areas, with the notable exception of thirteen and seventeen year olds' ability to compute with percentages. In the decade, both groups gained in that skill.

Computation by nine year olds provides the most dramatic contrast of the two decades. In the 1980s, nine year olds registered significant gains in the four operations of whole number arithmetic—addition, subtraction, multiplication, and division. In the 1990s, scores declined in all four operations. Changes in seventeen year olds' ability to compute with fractions also dramatically shifted over the two decades. In the 1980s, seventeen year olds' scores rose from 66.9% to 75.8%. Then in the 1990s proficiency plunged by more than twenty points to 55.5%.

Based on the 1999 scores, what generalizations can be made about how well American students compute?[14] If an 80% accuracy rate is considered indicating mastery, then nine year olds are close to mastering addition with whole numbers (75.5%) but still have not mastered subtraction (59.7%). They need considerable work before attaining mastery of multiplication (42.5%) or division (48.3%). At age thirteen, computing with whole numbers appears under control (89.5%), but computing with fractions and percentages remains challenging for most students (54.3% and 46.7%, respectively). And at age seventeen, computing with percentages shows promise (70.0%) but accuracy with fractions and decimals is weak (55.5% and 46.3%, respectively).

In sum, computation skills generally improved from 1982 to 1999. The improvement was almost entirely due to gains made in the 1980s. For the three age groups included in NAEP, the ability to compute accurately with whole numbers, decimals, and fractions declined in the 1990s. Computing with percentages was an exception, with thirteen and seventeen year olds showing marked improvement in the 1990s. Where do computation skills currently stand? Nine year olds appear to have mastered addition with whole numbers, but fall short of mastery in subtraction, multiplication, and division. Thirteen year olds appear to have mastered computing with whole numbers, but are still struggling with fractions and percentages. Seventeen year olds are within striking distance of demonstrating mastery in computing with percentages but not with fractions or decimals.

Race and Ethnic Gaps in Computation Skills

As noted earlier, the gap in NAEP mathematics test scores between white and black students reached its narrowest point in the late 1980s and then widened in the 1990s. A similar pattern is evident with computation skills (see Table 2).

Insert Tables 2, 3, and 3A

In 1982, the proficiency rates for white students surpassed those of black students in every skill area (see Table 2 for the scores of white students, Table 3 for blacks, and Table 3A for gaps). The gaps ranged from 6.6 percentage points for nine year olds on division (whites scoring 45.2% and blacks scoring 38.6%) to 25.0 percentage points for thirteen year olds computing with fractions (whites, 63.8% and blacks, 38.8%). All of the gaps shrank in the 1980s, with the exception of thirteen year olds' skills with percentages, and then reversed direction and expanded in the 1990s. The minimum gap in 1990 was 3.6 points; the maximum was 22.5. In 1999, the minimum gap was 7.9 points, and the maximum was 25.0.

Recall that computing with percentages is the skill area in which all thirteen and seventeen year olds made the greatest gains from 1982 to 1999. The national gains were mostly the result of increased proficiency by white students. Although black students registered gains, they were not as large. At age thirteen the black-white gap on percentages widened by 7.4 percentage points from 1982 to 1999, with most of the widening occurring in the 1990s. In 1999, whites outscored blacks by 20.6 points (whites, 50.9% and blacks, 30.3%). Among seventeen year olds, the black-white gap in computing with percentages narrowed in the 1980s and expanded in the 1990s. The gap stood at 25.0 percentage points in 1999, almost 4 points greater than it had been in 1982.

On the skills in which both white and black students suffered losses in the 1990s, the losses of black students were more severe. Thirteen year old white students' proficiency in computing with fractions declined about two percentage points (from 61.7% to 59.6%). Black students' proficiency fell nearly four points (from 39.2% to 35.2%), producing a 1.9 point widening of the gap. For both blacks and whites, seventeen year olds' ability to compute with fractions fell precipitously. For white students, scores declined by nearly 18 percentage points, from 76.8% to 59.0%. Black students' scores plummeted more than 33 percentage points, widening the black-white gap by 15.8 percentage points. It is important to note again that the relatively small African-American sample size increases the volatility of their scores; however, the t-statistics for the 1990s' changes exceeded 6.0 for all three fractions items so the decline is truly significant.

As the scorecard in Table 3B indicates, the 1990s were not a good time for reducing black-white gaps in computation skills. The 1980s were much more productive in that regard.

Insert Tables 4 and 4A

The Hispanic-white gaps in computation skills have not behaved in the same manner as the black-white gaps (see Tables 4, 4A, and 4B). The Hispanic-white gaps expanded in the 1980s and stayed relatively static in the 1990s. In comparing the 1980s and 1990s, the scorecard in Table 4B slightly favors the 1990s, with six positive trends and four negative ones.

Seventeen year olds' ability to compute with fractions is revealing. Like the black-white gap, the Hispanic-white gap contracted dramatically during the 1980s. White scores improved from 70.4% to 76.8%. Hispanic scores jumped from 43.5% to 63.2%, and the gap shrank by more than 23 percentage points. But unlike the black-white gap, the Hispanic-white gap held its own in the 1990s, remaining essentially unchanged. The news is not all good. The gap held steady because Hispanic and white skills declined the same amount, about 18 points.

A more consistently positive story is found with decimals. From 1982-1999, both Hispanics and whites improved their computing skills with decimals (whites by about nine points and Hispanics by about twenty-three), and the Hispanic-white gap shrank steadily through the 1980s and 1990s. Seventeen year olds' computation skills with percentages is a trouble spot. The Hispanic-white gap expanded during the 1980s by 4.9 percentage points and by 3.2 points during the 1990s. Similar to the expansion of the black-white gap in the same skill cluster, the Hispanic-white gap in computing with percentages throws cold water on an otherwise positive story. All students in 1999 knew how to compute with percentages better than in 1982, but gains by whites outstripped the gains of both blacks and Hispanics and, as a result, racial and ethnic gaps widened in this skill.

To sum up, the black-white and Hispanic-white gaps in computation skills did not act the same from 1982 to 1999. The black-white gap narrowed across the board in the

1980s, then reversed direction and grew wider in the 1990s. Most of the previous decade's progress was lost. The Hispanic-white gaps also showed significant narrowing in the 1980s, but unlike the black-white gaps, the Hispanic-white gaps held steady during the 1990s.

Computation Skills with Calculators Allowed: Matched Item Analysis

As noted above, the LTT NAEP has, on an experimental basis, included computation items that allow for the use of calculators.[15] The test for nine year olds has ten calculator items that also appear in the non-calculator portion of NAEP, allowing for a comparison of how students perform--on an identical set of problems--with and without calculators. The central questions are whether making calculators available influences conclusions about competency in a skill or trends in skill development, especially the trends examined above. Do we reach a different conclusion about how well nine year olds compute? Do we reach different conclusions about computation skills improving in the 1980s and declining in the 1990s or different conclusions about the performance gaps between whites and blacks or whites and Hispanics?

The ten items include three items in addition, three in subtraction, two in multiplication, and two in division.[16] It is important to stress that it is not known whether the nine year olds actually used calculators on these items, only that calculators were available for their use. References to "calculator items" in the following discussion should be interpreted accordingly. The results are presented in Table 5.

The total scores provide the headline. Large differences in performance occurred depending on whether students were allowed to use a calculator. In 1999, students averaged 85.3% accuracy on whole number computation items when calculators were allowed. On the same items, students only averaged 57.2% when calculators were not available and students were limited to pencil and paper for making calculations.

Insert Table 5

Digging deeper provides some important details. The differences appear in all skill clusters. Let's look at 1999 scores. With subtraction, students achieve 89.2% accuracy when calculators are available, but only 59.7% when calculators are withheld. With multiplication, accuracy reaches 87.9% with calculators, but only reaches 42.5% when calculators are not available. Proficiency with division reaches 77.1% with calculators and 48.3% without them. These differences are enormous--the difference between signaling mastery and signaling incompetence. Of the four operations, only addition appears close to mastery without a calculator (78.4%), consistent with the analysis presented above. These finding suggest that making calculators available on a test of computation skills can make the difference between concluding that students have acquired certain skills—and concluding that they haven't. On each of these items, at least 40% of the nation's nine year olds computed correctly with or without a calculator provided. For most of the remaining students, calculators are the difference in whether they compute correctly or get the calculation wrong.

Calculators also affect trends. Examine the two columns comparing changes in the two decades. Nine year olds made gains in the 1980s--whether calculators were allowed or not . However, they made greater gains when calculators were withheld (7.0

percentage points for the total score) than when present (2.1 percentage points). In the 1990s, student performance improved when calculators were allowed (2.4 percentage points), but declined when calculators were unavailable (-1.9 percentage points). This pattern—gains with calculators and losses without them—occurred in all four operations in the 1990s. In addition to affecting our perception of students' skill at computing with whole numbers, then, calculators also can affect whether changes over time are judged to be positive or negative. In the 1990s, nine year olds' computation scores rose from 83% to 85% when calculators were available on the test items. Skills were solid and improving. Given the same items, but without the aid of calculators, student performance slipped from 59% to 57%. Skills were weak and getting weaker.

Insert Tables 6, 7, and 7A here

Tables 6-7A show black, white, and performance gap scores on the matched items. Again we see that calculator availability shapes the appearance of trends. From 1982-1999, blacks gained more when calculators were available than when they were not, 10.7 vs. 5.9 percentage points. Black-white gaps narrowed in most skill areas from 1982-1999 (see Table 7A). In every case, the narrowing was more pronounced when calculators were present than when they were withheld. Trends in the 1990s are particularly disparate due to the calculator effect. Again, total scores tell the major story. The black-white proficiency gap widened in the 1990s by 3.1 percentage points on pencil and paper computation skills. With calculators available, the gap moved in the opposite direction, narrowing by 4.1 points. In terms of student performance on these ten computation items, greater equity can appear to be achieved by changing a single test condition—by shifting from paper and pencil to allowing calculators.

Insert Table 8 and 8A here

Scores for Hispanic students are reported in Table 8. The Hispanic-white gaps are presented in Table 8A. From 1982-1990, the same effect surfaces that was detected with the black-white gaps. With calculators allowed, considerable progress was made. The Hispanic-white gap in whole number computation contracted by 6.9 percentage points. Progress evaporated when calculators were taken away. The gap expanded by 0.8 of a point.

Discussion

The study analyzed data from computation items on the LTT NAEP assessment. A subset of matched items compared how well nine year olds compute when calculators are available and when they are not. The items were not designed to assess computation skills comprehensively, and, therefore, the findings should be considered as suggestive rather than conclusive. Despite this limitation, the items represent the only set of computation items administered periodically to a large, randomly selected national sample of students.

Six research questions were presented at the beginning of the paper. The following discussion is organized around the findings on each question.

Did computation skills improve, decline, or remain the same from 1982 to 1999? Computation skills generally improved from 1982 to 1999. Skills of nine year olds increased in all skill areas examined, and thirteen year olds and seventeen year olds skills improved in percentages only. Thirteen year olds' skills fell for whole number and fractions, and seventeen year olds struggled more with decimals and fractions.

How do the 1980s and 1990s compare on trends in computation skills? The 1990s were not as productive as the 1980s in boosting computation skills. Virtually all of the gains from 1982 to 1999 occurred in the 1980s and then faded in the 1990s. Nine year olds' skills declined in adding subtracting, multiplying, and dividing whole numbers. Thirteen year olds' skills declined in computing with whole numbers and fractions. Seventeen year olds' skills declined in computing with fractions and decimals. The single bright spot of the 1990s was in computing with percentages, with both thirteen and seventeen year olds showing considerable improvement.

Why did computation skills reverse direction and decline in the 1990s? There are several plausible explanations. Math reform in the 1990s, which the NCTM was influential in promoting but not the sole advocate of, de-emphasized computation skills in favor of problem solving, geometry, data and statistical analysis, and other areas of mathematics. New textbooks reflected these priorities. Use of calculators increased significantly in classrooms during the decade, even in the fourth grade.[17] Research has not yet determined whether calculators affect children's learning how to compute in fourth grade and lower. Although no one can be certain what caused the decline in computation, aspects of math reform in the 1990s are reasonable candidates for at least contributing to the fall off in skills.

How sound were students computational skills as of 1999, the last time the trend NAEP was administered? In 1999, nine year olds showed evidence of nearly mastering addition with whole numbers (75.5%). Subtraction remained a problem area (59.7%), and mastery of multiplication (42.5%) and division (48.3%) was even further out of reach. The nation is not even close to meeting the goal of the NCTM's 2000 standards—that all students should be fluent in whole number computation by the end of fifth grade. Thirteen year olds evidence mastery of whole numbers (89.5%), but have difficulty computing with fractions (54.3%) or percentages (46.7%). Seventeen year olds are fairly proficient computing with percentages (70.0%) but have difficulty with fractions (55.5%) and decimals (46.3%).

In 1999, how large were black-white and Hispanic-white gaps in computation skills? Have they changed over time? Trends in black-white gaps in computation skills mirror achievement gaps in mathematics more generally—significant narrowing in the 1980s, followed by modest widening in the 1990s. In the 1980s both groups gained in skill, but blacks gained more than whites. In the 1990s, both groups' computation skills declined, but the skills of blacks declined more. In terms of absolute level of performance in individual skill areas, whites outperformed blacks by eight to twenty-five percentage points in 1999. At age thirteen, for example, the cluster with the narrowest gap is whole number computation. Whites scored 90.8%, and blacks scored 82.9%. The widest gap was registered with fractions. Whites scored 59.6%, and blacks scored 35.2%.

Trends in the Hispanic-white gap are not as consistent. Like black and white students, Hispanic students improved in the ability to compute in the 1980s. But in the 1990s Hispanic scores did not fall as much as the scores of blacks. The gaps with white students held steady. The Hispanic-white gaps also are not as large as those between black and white students. In individual skill areas, whites outperformed Hispanics by approximately two to seventeen percentage points in 1999. The narrowest gap was discovered in nine year olds' ability to divide whole numbers. Hispanic students scored 47.7%, and white students scored 49.4%. The largest gap was in seventeen year olds' ability to compute with percentages. Hispanics scored 57.9%, and whites scored 75.4%.

Given an identical set of computation items, how do nine year olds perform when they are allowed to use a calculator compared to when a calculator is not allowed? Calculators change everything. For a large number of nine year olds, when calculators are available on computation items, they get correct answers. When calculators are not available, they get wrong answers.   The smallest calculator advantage is on addition items, a skill nearing mastery. Students score 78.4% using only pencil and paper and 87.0% with calculators.   The calculator advantage in other areas is huge. In subtraction, students scored 89.2% with calculators available and 59.7% without calculators. In multiplication, students scored 87.9 % with calculators and 42.5% without them.   On division items, the scores were 77.1% with calculators, 48.3% without.

The conclusion is clear: allowing fourth graders to use calculators on items that are intended to assess computation skills will produce misleading results—misleading, that is, if one assumes that knowing how to compute means being able to make calculations without technological assistance. The fear of critics that calculators might serve as a crutch also appear well founded. Believing that a nine year old can compute when he or she cannot do so without a calculator is tantamount to believing that a nine year old can ride a bike when he or she cannot do so without training wheels.

Do the conclusions about the trends analyzed above (comparison of the 1980s and 1990s, progress in closing black-white and Hispanic gaps) differ when calculators are available or not? Calculators not only affect absolute levels of computation scores, but also trends. The 1990s decline in computation scores that occurred when nine year olds were limited to pencil and paper did not occur when students had a calculator available. Scores actually increased. Why would this happen? The most likely explanation is that, as calculator usage increased in the 1990s, students became more proficient at using them. The difference between students' accuracy with paper and pencil calculations and calculations with a calculator widened.

Why would racial and ethnic gaps in computation skills also be affected? For blacks and Hispanics, the performance differences between calculator and no calculator test conditions are much wider than for whites. Calculators make a bigger difference with students of color. There are two possible explanations. The first is racial differences in calculator usage. In 1996, black and Hispanic fourth graders in the NAEP survey were about twice as likely as white students to report using calculators "everyday" for math class work.[18] But those disparities evaporated in the 2000 and 2003 surveys, which were administered after the 1999 endpoint for data in the current study. White fourth graders reported using calculators more often. So that explanation may no longer hold water. Another explanation pertains to skill levels themselves. If black and Hispanic computation skills have indeed slipped more than whites, then that alone could make the calculator advantage grow disproportionately. The weakness of that explanation is that Hispanics' computation skills did not decline in the 1990s, yet the gap between using and not using calculators grew anyway. This topic needs much more research to fully understand racial and ethnic differences in the calculator advantage

Conclusion

Let's conclude with a key question. Why do the results of this study matter? As pointed out above, the main NAEP and most state tests allow students to use calculators on at least a portion of the exams. The reason usually offered is that problem solving tasks are designed to measure the ability to solve problems, not the ability to compute. That makes sense, but the findings presented here present a challenge. Faced with virtually any mathematical problem that is worth solving, students must compute an answer. If students are only able to compute accurately with calculators—or if their computation skills are so weak that only the simplest of calculations can be made—then students are doomed to solving only trivial mathematical problems.

Another important consideration involves the need for all of us—parents, teachers, taxpayers, and the public—to find out what students have learned and what they still need to learn. In contrast to the decline in computation skills documented here, the main

NAEP and state tests have been showing tremendous growth in students' mathematical achievement. These tests do not report a computation score. For the most part, they do not include many computation items, even when assessing children in fourth grade. Officials at both the federal and state levels should reconsider these policies—for the sake of children's learning and for the public's right to know.

---

[1] See Tom Loveless, "The Regulation of Teaching and Learning," in *Stability and Change in American Education: Structure, Process, and Outcomes*, edited by Maureen Hallinan, Adam Gamoran, Warren Kubitschek, and Tom Loveless, (Eliot Werner Publications, Inc., 2003), pp. 171-192.

[2] David Klein and R. James Milgram, "The Role of Long Division in the K-12 Curriculum," February 2000; H. Wu, "How to Prepare Students for Algebra," *American Educator* (Summer 2001), pp. 1-7. For an influential parent group on mathematics, see www.mathematicallycorrect.com.

[3] National Council of Teachers of Mathematics (NCTM), *Principles & Standards for School Mathematics*, (NCTM, 2000), p. 152.

[4] See the NAEP website's acknowledgement of the NCTM's influence on the math framework at: http://nces.ed.gov/nationsreportcard/mathematics/whatmeasure.asp

[5] Divergences from NCTM do exist. For example, NCTM has supported allowing calculators on more items than currently permitted. But--make no mistake about it--the NCTM's influence runs throughout the frameworks adopted in the 1990s. The College Board organized the Mathematics Consensus Project, the group responsible for drafting the 1996 framework. Its report recommending several changes to NAEP ends with this assurance; "The framework moves NAEP closer to the ideal described in the NCTM standards (*Mathematics Framework for the 1996 National Assessment of Educational Progress*, page 7). " John Dossey, president of NCTM when the NCTM standards were developed, served as director of the Mathematics Consensus Project.

[6] In 2004, Minnesota added seven computation items to the state's basic skills test and will not allow calculators on them. John Welsh, "Calculators Now Off Limits for Part of Basic Skills Test," *Pioneer Press*, January 4, 2004. Available at www.twincities.com/mld/pioneerpress.

[7] For criticism of NSF's activities in mathematics, see Michael McKeown, David Klein, and Chris Patterson, "National Science Foundation Systematic Initiatives," in *What's at Stake in the K-12 Standards Wars - A Primer for Educational Policy Makers*, edited by Sandra Stotsky (Peter Lang, 2000), pp. 313-369. On the NSF's support for NCTM-oriented textbooks, see Mark Clayton, "How a New Math Program Rose to the Top," *Christian Science Monitor*, May 23, 2000, p. 15.

[8] David Klein, "Should We Curb the Use of Calculators?" *American Educator* (March, 2001).

[9] Public Agenda Foundation, *Different Drummers: How Teachers of Teachers View Public Education*, (Public Agenda Foundation, 1997).

[10] Aimee J. Ellington, "Effects of Calculators: A Meta-Analysis," *Journal for Research in Mathematics Education*, vol. 34, no. 5 (November 2003), pp. 433-463. Quote from page 435. Ray Hembree and Donald J. Dessart, "Efffects of Hand-Held Calculators in Pre-College Mathematics Education: A Meta-Analysis," *Journal for Research in Mathematics Education*, vol. 17 (1986), pp. 83-99; Brian A. Smith, *A Meta-Analysis of Outcomes from the Use of Calculators in Mathematics Education*. Ph.D. dissertation, College of Education, Texas A&M University (1996).

[11] Jaekyung Lee, "Racial and Ethnic Achievement Gap Trends: Reversing the Progress Toward Equity," *Educational Researcher*, vol. 31. no 1, pp 3-12 (January/February 2002). Also see Larry V. Hedges and Amy Nowell, "Black-White Test Score Convergence Since 1965," pp. 149-181 in *The Black-White Test Score Gap*, edited by C. Jencks and M. Phillips (Washington DC: Brookings Institution Press, 1998).

[12] U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, *NAEP 1999 Trends in Academic Progress: Three Decades of Student Performance*, by J.R. Campbell, C.M. Hombo, and J. Mazzeo, (NCES 2000?469, U.S. Government Printing Office, 2000).

[13] Tom Loveless, *The 2002 Brown Center Report on American Education*, (Brookings Institution Press, 2002), pp. 6-14.

[14] The long term trend NAEP is currently being given in the 2003-04 academic year.

[15] The main NAEP allows calculator use on 35% to 40% of items and includes them in all scores. Also, contrary to the long term trend test, on most questions with calculator availability the main NAEP asks students whether they used a calculator to solve the problem.

[16] A few matched items were available at age thirteen. We found no significant difference in student performance on computation items involving percentages whether calculators were used or not.

[17] Data on calculator use obtained by using the NAEP Data Tool at: http://nces.ed.gov/nationsreportcard/naepdata/search.asp

[18] Also see Daniel Golden, "For Inner-City Schools, Calculators May Be The Wrong Answer," *The Wall Street Journal*, December 15, 2000.

Table 1. NAEP Computation Items

No Calculator allowed
All Students

	Skill Cluster (Items)	1982	1990	1999	Change 1982-1990	Change 1990-1999	Change 1982-1999
Age 9	**Addition (4)**	71.4	75.7	75.5	4.3	-0.2	4.1
	Subtraction (3)	53.4	63.4	59.7	10.0	-3.7	6.3
	Multiplication (2)	37.4	43.9	42.5	6.5	-1.4	5.1
	Division (2)	44.3	49.1	48.3	4.8	-0.8	4.0
Age 13	**Whole Numbers (9)**	92.7	91.4	89.5	-1.3	-1.9	-3.2
	Fractions (4)	59.7	56.8	54.3	-2.9	-2.5	-5.4
	Percentages (6)	39.1	38.8	46.7	-0.3	7.9	7.6
Age 17	**Fractions (3)**	66.9	75.8	55.5	8.9	-20.3	-11.4
	Decimals (3)	51.3	50.4	46.3	-0.9	-4.1	-5.0
	Percentages (6)	56.3	66.5	70.0	10.2	3.5	13.7

Table 1a. How do the 1990s compare with the 1980s?

Positive Trend Changes in 1990s	
Losses Reversed	1
Gains Accelerated	0
Losses Slowed	0
Total	1

Negative Trend Changes in 1990s	
Gains Reversed	5
Losses Accelerated	3
Gains Slowed	1
Total	9

Table 2. NAEP Computation Items

No Calculator Allowed

White Students

	Skill Cluster (Items)	1982	1990	1999	Change 1982-1990	Change 1990-1999	Change 1982-1999
Age 9	Addition (4)	73.7	77.6	77.4	3.9	-0.2	3.7
	Subtraction (3)	55.8	65.8	63.4	10.0	-2.4	7.6
	Multiplication (2)	38.4	44.7	43.5	6.3	-1.2	5.1
	Division (2)	45.2	49.5	49.4	4.3	-0.1	4.2
Age 13	Whole Numbers (9)	93.9	92.4	90.8	-1.5	-1.6	-3.1
	Fractions (4)	63.8	61.7	59.6	-2.1	-2.1	-4.2
	Percentages (6)	41.2	41.9	50.9	0.7	9.0	9.7
Age 17	Fractions (3)	70.4	76.8	59.0	6.4	-17.8	-11.4
	Decimals (3)	71.8	83.1	81.2	11.3	-1.9	9.4
	Percentages (6)	59.3	69.5	75.4	10.2	5.9	16.1

Table 3. NAEP Computation Items

No Calculator Allowed

Black Students

	Skill Cluster (Items)	1982	1990	1999	Change 1982-1990	Change 1990-1999	Change 1982-1999
Age 9	Addition (4)	59.1	68.4	67.9	9.3	-0.5	8.8
	Subtraction (3)	40.3	53.5	45.4	13.2	-8.1	5.1
	Multiplication (2)	28.8	38.8	35.2	10.0	-3.6	6.4
	Division (2)	38.6	45.9	39.9	7.3	-6.0	1.3
Age 13	Whole Numbers (9)	85.5	87.6	82.9	2.1	-4.7	-2.6
	Fractions (4)	38.8	39.2	35.2	0.4	-4.0	-3.6
	Percentages (6)	28.0	26.5	30.3	-1.5	3.8	2.3
Age 17	Fractions (3)	50.6	73.2	39.6	22.6	-33.6	-11.0
	Decimals (3)	51.8	67.7	65.2	15.9	-2.5	13.4
	Percentages (6)	38.2	53.8	50.4	15.6	-3.4	12.2

Table 3a. NAEP Computation Items
No Calculator Allowed
Black-White Gap

	Skill Cluster (Items)	1982	1990	1999	Change 1982-1990	Change 1990-1999	Change 1982-1999
Age 9	**Addition (4)**	14.6	9.2	9.5	-5.4	0.3	-5.1
	Subtraction (3)	15.5	12.3	18.0	-3.2	5.7	2.5
	Multiplication (2)	9.6	5.9	8.3	-3.7	2.4	-1.3
	Division (2)	6.6	3.6	9.5	-3.0	5.9	2.9
Age 13	**Whole Numbers (9)**	8.4	4.8	7.9	-3.6	3.1	-0.5
	Fractions (4)	25.0	22.5	24.4	-2.5	1.9	-0.6
	Percentages (6)	13.2	15.4	20.6	2.2	5.2	7.4
Age 17	**Fractions (3)**	19.8	3.6	19.4	-16.2	15.8	-0.4
	Decimals (3)	20.0	15.4	16.0	-4.6	0.6	-4.0
	Percentages (6)	21.1	15.7	25.0	-5.4	9.3	3.9

Table 3b. How do the 1990s compare with the 1980s?

Positive Trend Changes in 1990s	
Widening Reversed	0
Narrowing Accelerated	0
Widening Slowed	0
Total	0

Negative Trend Changes in 1990s	
Narrowing Reversed	9
Widening Accelerated	1
Narrowing Slowed	0
Total	10

Table 4. NAEP Computation Items
No Calculator Allowed
Hispanic Students

	Skill Cluster (Items)	1982	1990	1999	Change 1982-1990	Change 1990-1999	Change 1982-1999
Age 9	Addition (4)	67.9	69.3	69.5	1.4	0.2	1.6
	Subtraction (3)	44.3	53.9	51.3	9.6	-2.6	7.0
	Multiplication (2)	40.5	41.0	40.8	0.5	-0.2	0.3
	Division (2)	42.8	43.0	47.7	0.2	4.7	4.9
Age 13	Whole Numbers (9)	91.7	88.5	88.1	-3.2	-0.4	-3.6
	Fractions (4)	49.7	45.6	42.8	-4.1	-2.8	-6.9
	Percentages (6)	35.9	33.8	38.5	-2.1	4.7	2.6
Age 17	Fractions (3)	43.5	63.2	45.2	19.7	-18.0	1.7
	Decimals (3)	49.4	69.7	72.8	20.3	3.1	23.4
	Percentages (6)	49.9	55.2	57.9	5.3	2.7	8.0

Table 4a. NAEP Computation Items

No Calculator Allowed

Hispanic-White Gap

	Skill Cluster (Items)	1982	1990	1999	Change 1982-1990	Change 1990-1999	Change 1982-1999
Age 9	**Addition (4)**	5.8	8.3	7.9	2.5	-0.4	2.1
	Subtraction (3)	11.5	11.9	12.1	0.4	0.2	0.6
	Multiplication (2)	2.1	3.7	2.7	1.6	-1.0	0.6
	Division (2)	2.4	6.5	1.7	4.1	-4.8	-0.7
Age 13	**Whole Numbers (9)**	2.2	3.9	2.7	1.7	-1.2	0.5
	Fractions (4)	14.1	16.1	16.8	2.0	0.7	2.7
	Percentages (6)	5.3	8.1	12.4	2.8	4.3	7.1
Age 17	**Fractions (3)**	36.9	13.6	13.8	-23.3	0.2	-23.1
	Decimals (3)	22.4	13.4	8.4	-10.0	-5.0	-15.0
	Percentages (6)	9.4	14.3	17.5	4.9	3.2	8.1

Table 4b. How do the 1990s compare with the 1980s?

Positive Trend Changes in 1990s	
Widening Reversed	4
Narrowing Accelerated	0
Widening Slowed	2
Total	6

Negative Trend Changes in 1990s	
Narrowing Reversed	1
Widening Accelerated	2
Narrowing Slowed	1
Total	4

Table 5. Matched Calculator Items

All Students

	Skill Cluster (Items)	Calc?	1982	1990	1999	Change 1982-1990	Change 1990-1999	Change 1982-1999
Age 9 Whole Number Operations	Add (3)	N	73.3	80.0	78.4	6.7	-1.6	5.1
		Y	79.5	83.2	87.0	3.7	3.8	7.5
	Subtract (3)	N	53.4	63.4	59.7	10.0	-3.7	6.3
		Y	87.0	88.3	89.2	1.3	0.9	2.2
	Multiply (2)	N	37.4	43.9	42.5	6.5	-1.4	5.1
		Y	83.9	86.8	87.9	2.9	1.1	4.0
	Divide (2)	N	44.3	49.1	48.3	4.8	-0.8	4.0
		Y	72.8	73.3	77.1	0.5	3.8	4.3
	Total (10)	N	52.1	59.1	57.2	7.0	-1.9	5.1
		Y	80.8	82.9	85.3	2.1	2.4	4.5

Table 6. Matched Calculator Items

White Students

	Skill Cluster (Items)	Calc?	1982	1990	1999	Change 1982-1990	Change 1990-1999	Change 1982-1999
Age 9 Whole Number Operations	Add (3)	N	75.5	81.8	80.1	6.3	-1.7	4.6
		Y	82.6	84.9	88.6	2.3	3.7	6.0
	Subtract (3)	N	55.8	65.8	63.4	10.0	-2.4	7.6
		Y	89.9	90.7	90.4	0.8	-0.3	0.5
	Multiply (2)	N	38.4	44.7	43.5	6.3	-1.2	5.1
		Y	86.5	88.9	89.0	2.4	0.1	2.5
	Divide (2)	N	45.2	49.5	49.4	4.3	-0.1	4.2
		Y	74.9	75.7	79.0	0.8	3.3	4.1
	Total (10)	N	56.2	63.1	61.6	6.9	-1.5	5.4
		Y	76.4	77.8	79.3	1.4	1.5	2.9

Table 7. Matched Calculator Items

Black Students

	Skill Cluster (Items)	Calc?	1982	1990	1999	Change 1982-1990	Change 1990-1999	Change 1982-1999
Age 9 Whole Number Operations	**Add (3)**	N	65.5	72.7	71.9	7.2	-0.8	6.4
		Y	66.5	74.2	80.7	7.7	6.5	14.2
	Subtract (3)	N	40.3	53.5	45.4	13.2	-8.1	5.1
		Y	72.5	77.6	84.1	5.1	6.5	11.6
	Multiply (2)	N	28.8	38.8	35.5	10.0	-3.3	6.7
		Y	71.2	78.1	83.2	6.9	5.1	12.0
	Divide (2)	N	38.6	45.9	39.9	7.3	-6.0	1.3
		Y	60.3	61.5	68.3	1.2	6.8	8.0
	Total (10)	N	44.3	54.8	50.2	10.5	-4.6	5.9
		Y	61.8	66.8	72.5	5.0	5.7	10.7

Table 7a. Matched Calculator Items

Black-White Gap

	Skill Cluster (Items)	Calc?	1982	1990	1999	Change 1982-1990	Change 1990-1999	Change 1982-1999
Age 9 Whole Number Operations	**Add (3)**	N	13.0	9.1	8.2	-3.9	-0.9	-4.8
		Y	16.0	10.7	7.9	-5.3	-2.8	-8.1
	Subtract (3)	N	15.5	12.3	18.0	-3.2	5.7	2.5
		Y	17.3	13.1	6.3	-4.2	-6.8	-11.0
	Multiply (2)	N	9.7	5.9	8.4	-3.8	2.5	-1.3
		Y	15.3	10.8	5.8	-4.5	-5.0	-9.5
	Divide (2)	N	6.6	3.6	9.6	-3.0	6.0	3.0
		Y	14.6	14.2	10.7	-0.4	-3.5	-3.9
	Total (10)	N	11.8	8.3	11.4	-3.5	3.1	-0.4
		Y	14.5	11.0	6.9	-3.5	-4.1	-7.6

Table 8. Matched Calculator Items

Hispanic Students

		Skill Cluster (Items)	Calc?	1982	1990	1999	Change 1982-1990	Change 1990-1999	Change 1982-1999
Age 9	Whole Number Operations	Add (3)	N	68.9	73.5	74.1	4.6	0.6	5.2
			Y	67.6	82.1	85.9	14.5	3.8	18.3
		Subtract (3)	N	44.3	54.0	51.3	9.7	-2.7	7.0
			Y	82.9	82.6	88.7	-0.3	6.1	5.8
		Multiply (2)	N	40.5	41.0	40.8	0.5	-0.2	0.3
			Y	76.4	83.4	87.7	7.0	4.3	11.3
		Divide (2)	N	42.8	43.0	47.7	0.2	4.7	4.9
			Y	70.1	72.3	76.6	2.2	4.3	6.5
		Total (10)	N	50.6	55.0	55.3	4.4	0.3	4.7
			Y	67.7	73.3	77.5	5.6	4.2	9.8

Table 8a. Matched Calculator Items

Hispanic-White Gap

		Skill Cluster (Items)	Calc?	1982	1990	1999	Change 1982-1990	Change 1990-1999	Change 1982-1999
Age 9	Whole Number Operations	Addition (3)	N	6.6	8.3	5.9	1.7	-2.4	-0.7
			Y	15.0	2.9	2.6	-12.1	-0.3	-12.4
		Subtract (3)	N	11.5	11.9	12.1	0.4	0.2	0.6
			Y	6.9	8.1	1.6	1.2	-6.5	-5.3
		Multiply (2)	N	-2.1	3.7	2.7	5.8	-1.0	4.8
			Y	10.1	5.4	1.3	-4.7	-4.1	-8.8
		Divide (2)	N	2.4	6.5	1.7	4.1	-4.8	-0.7
			Y	4.8	3.4	2.5	-1.4	-0.9	-2.3
		Total (10)	N	5.5	8.1	6.3	2.6	-1.8	0.8
			Y	8.7	4.6	1.8	-4.1	-2.8	-6.9

Appendix A. Significance Tests of Changes in Matched Computation Items for 9-year olds

		1982-1990		1990-1999	
	Sig. Level	Calc	No Calc	Calc	No Calc
All Students	p<.05	2-0-8	8-0-2	2-0-8	0-0-10
	p<.10	2-0-8	8-0-2	4-0-6	0-2-8
White Students	p<.05	2-0-8	8-0-2	2-0-8	0-0-10
	p<.10	2-0-8	8-0-2	3-0-7	0-1-9
Black Students	p<.05	3-0-7	9-0-1	3-0-7	0-1-9
	p<.10	4-0-6	9-0-1	4-0-6	0-3-7
Hispanic Students	p<.05	3-0-7	1-0-9	1-0-9	1-0-9
	p<.10	4-1-5	1-0-9	1-0-9	2-0-8

Note: First number in each cell represents number of items on which there were statistically significant positive changes. The second number is the number of statistically significant negative changes. The third number shows the number of items on which the change was not statistically different than zero.